# AN ECONOMIC ANALYSIS OF REGULATORY OBJECTIVES AND ENFORCEMENT

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Doctor of Philosophy

in the

University of Canterbury

by

David J. Walker

University of Canterbury

1989

# CONTENTS

## LIST OF FIGURES

## ABSTRACT

Theories of regulation are many. No single theory however explains the observed contrasts in regulatory behaviour and no clear criteria exist by which to identify a particular regulation as according to one theory or another. This is because there is no consistent theoretical structure which links the various approaches. Such a framework is constructed here.

A partial equilibrium model of regulation in a competitive, negative externality generating industry is formulated and the concept of a regulated equilibrium defined. Regulatory controls over economic behaviour are ineffective without enforcement. In the absence of enforcement the regulated equilibrium coincides with the unregulated competitive equilibrium. Enforcement is by means of an expected monetary penalty. The effects of changes in policy instruments on the regulated equilibrium are derived.

Within this otherwise identical framework two competing hypotheses concerning the aims of regulation are introduced, namely the Naive Public Interest Approach and the Capture Theory. Each objective is defined within the regulatory environment realizing that enforcement is costly. Optimal regulatory policy under each hypothesis is determined and policy responses to parameter changes are analysed.

Most theories of economic regulation are set in a partial equilibrium framework. The implications for

regulation and enforcement of the regulation's impact on sectors of an economy other than that which is directly affected are therefore not considered. These ignored sectors may, however, affect the outcome. In light of this the present model is reformulated in a simple general equilibrium context. Resource requirements for enforcement are explicitly considered in the derivation of the economy's regulated equilibrium locus of which the optimal equilibria under the competing regulatory hypotheses form a part. The behavioural characteristics of these policies are derived and compared. It is shown that potentially observable differences in regulatory policy exist from which the regulator's objectives can be inferred.

# CHAPTER ONE

# INTRODUCTION AND LITERATURE REVIEW

## 1-1 INTRODUCTION

Throughout history, man has organised himself into societal groupings in which the individual is subject to a central governing authority. The form of government, and the nature and extent of its powers, varies between different societies and within particular societies over time. The delineation of the central authority's powers and the extent of an individual's obligation to it have long been the topic of debate. This has been so from the writings of early philosophers such as Plato and Aristotle, through to Hobbes, Locke and Hume in seventeenth and eighteenth century England, and to the present day.

One of the most common and also most contentious areas of government involvement continues to be in the economic sphere of life. The economic activities of a government can range from the provision of simple legal structures which facilitate private commerce to detailed controls over individuals' economic actions and explicit state ownership and operation of commercial ventures.

The essentially liberal writings of Locke laid the foundation for the libertarian, laissez-faire nature of modern economic thought espoused initially by Hume and enshrined in Smith's notion of the "invisible hand". This intellectual trend continued largely unchallenged in the mainstream literature for the next century, culminating in the writings of Marshall.

Perhaps because of its pre-eminent position, a degree of complacency pervaded the liberal school. This complacency has been berated by Stigler.

> So much for the century of laissez-faire. The main school of economic individualism had not produced even a respectable modicum of evidence that the state was incompetent to deal with economic problems of any or all sorts ... the doctrine of non-intervention was powerful only so long and so far as man wished to obey. [Stigler, 1975; 45]

Beginning in England from the mid-nineteenth century onwards, and following slightly later in the United States, a greater degree of public control began to be exercised over economic life. Demands for increased state intervention followed from the identification of what are now called "market failures" caused by factors such as externalities, public goods, and informational asymmetries, and a concern that insufficient information often led to individual agents making incorrect decisions with respect to time preference.

These arguments were strengthened by the massive supposed market failure that was the Great Depression in the years 1929-1933.[1] The response in the United States was the 'New Deal' [Friedman and Schwartz, 1963; 420] which incorporated a much greater involvement of the state in economic affairs than had previously been the case, and paved the way for the explosion of detailed regulatory activity that has subsequently occurred.

It is important to remember that, for the most part, markets don't fail. Indeed, western-style economies are based on the belief that the market mechanism is, in general, the most efficient method of determining resource allocation. Nevertheless, economists' attentions are naturally drawn to the pathological cases where they do fail. Concentration on such aberrations tends to create an impression that market failure is the norm.

As the welfare economists espoused their views on the failures of the market, there was a concomitant belief in the omnipotence of the state to correct these misallocations. Regulations were enacted to protect the public at large from the evils of others and even from themselves. It was argued that consumers needed to be protected from shoddy merchants and monopolists, and that producers required a degree of protection from the vagaries of unstable market conditions in order to ensure continuous and efficient service to their customers. These arguments were apparently successful in initiating many regulatory institutions [Owen and Braeutigam, 1978; 10].

This view held that regulation served the public interest. With the passage of time, however, it became increasingly obvious that regulation was not fulfilling its imagined role as public protector. Regulations often appeared at best ineffective in restraining monopoly, and at worst facilitated restrictive practices through processes akin to cartel management. In the 1960's and early 1970's, the role of regulation as a promoter of the public interest began to be questioned in the professional literature[2] as a search began for a theory to explain the regulatory phenomenon.

Proposed theories of economic regulation are many and varied. The next two sections of the Chapter briefly outline some of the major theories concerning the existence of regulation and its effect on the regulated industry.

Section 1-4 introduces an aspect of regulation that was ignored in the early literature and has received only relatively recent attention; namely the necessity for enforcement. This aspect of the regulatory problem is developed and a summary of the analysis undertaken in the remaining chapters of the Thesis is presented.

## 1-2 THEORIES OF REGULATION

Stigler defines regulation as

> ... an attempt by the state to use its legal powers to direct the conduct ... of nongovernmental bodies. [Stigler, 1981; 73]

He goes on to say that this implies that public regulation includes

> ... most of public finance, large parts of monetary and financial economics and international trade, large sectors of labour economics, agricultural and land economics, and welfare economics. [Stigler, 1981; 73]

The scope of economic regulation would appear, from this definition, to be extremely broad and indeed the theories and hypotheses advanced to explain its existence are too numerous to attempt an exhaustive coverage here. Some of the main approaches however are presented below.

### (1) Public Interest

The traditional approach to regulation was based on the assumption that it facilitated the public interest. This now-called "naive public interest" view held that regulation was supplied in response to public demand for the correction of inefficient or inequitable market practices. It was based on the dual belief that markets, if left alone were inefficient and inequitable, and that the panacea for these ills was government regulation which was essentially costless. This implies that the supply and demand for regulation itself is somehow immune from market failure, an implication which forms the basis of much of the subsequent attacks on the public interest theory.

Posner [1974] suggests that if the public interest hypothesis was a true characterization of the regulatory process, regulations would be imposed mainly in highly concentrated industries where the threat of monopoly exists, and that substantial external costs or benefits would be generated. He then asserts that neither of these characteristics is observable in practice. He goes on to say that several attempts have been made to reformulate the naive public interest theory into a more empirically relevant explanation of the regulatory process. In these, regulatory agencies are viewed as being created for legitimate concerns, but then are either mismanaged or engage in honest but unsuccessful efforts to promote the public interest.[3] The efforts are unsuccessful because firstly, many of the tasks an agency is required to perform, such as industry cost determination, are extremely difficult, and secondly, transactions costs deter sufficient effective legislative supervision of agencies' performance [Posner, 1974; 337-340].

One of the major problems with these theories is that they contain no clear linkage mechanism whereby perception of the public interest is translated into legislative action. Presumably in a democratic system, this must occur in some manner through the political system. It is precisely this political process however that other writers have concentrated on to show that regulation delivered in a democratic system does not primarily serve the public interest.[4] This will be further discussed later in the section when private interest theories are introduced.

(2) Regulation as Contract

Goldberg [1976] draws an analogy between regulation and long term contracts. He analyses the features of long term contracts between producers and consumers in which the parties to the agreement voluntarily limit their future options in order to minimize uncertainty and costs of adjustment. He suggests that there are strong similarities between these contractural arrangements and the process of economic regulation and that the regulatory body may be viewed as an agent of a consumer group in negotiating and administering a contract.

It is possible using this theory to explain restriction of entry in decreasing cost, natural monopoly industries, such as public utilities, as a rational response by consumers to avoid the over-capitalization that could initially occur under competition. The theory however does not easily accord with regulation of essentially competitive industries unless, in this case, the regulator is acting as an agent of the producers in order to achieve positive industry profits. The most likely role of the regulator in this framework then is that of an arbiter, not unlike a judge in court, who considers the advocacies of both sides in determining regulatory policy. The theory then becomes similar to that of Peltzman [1976] and Becker [1983] where the arbiter is the political process.

Owen and Braeutigam [1978; 18-32] also concentrate on the legal aspects of regulation. They stress the role of law in providing economic justice or fairness[5] as the objective of, and motivating force behind, the creation of regulations. The two features of the legal system that they emphasize are delay, and "derivative": the granting of equity rights to individuals or groups in the status quo [Owen and Braeutigam, 1978; 20].

> ... the effect of administrative procedure, the legal rules that constrain the forms of regulatory decision making, is to slow down or delay the operation of market forces. [Owen and Braeutigam, 1978; 18-19]

They argue that the contrast between the ideological principle that markets should be free from the incompetence of government, and the fact that regulation has rapidly increased over the last fifty years, can be explained by examining the preferences and incentives of voters.[6]

Voters, they suggest, may be expected to be risk averse whereas the free market is a risky institution. In a free market, the only vote an agent has is his purchasing power. Regulation in effect enshrines a legal right to prices, quality of goods and services, and market share. All of these are easily disturbed in the free market, often adversely for many agents. The effect of the administrative process is to increase the voting power of the harmed agents by granting them due process. Compensation of some kind is often granted in the form of subsidy or time delay to facilitate adjustment, and, if the compensation is insufficient, the change can often be effectively blocked.[7]

The administrative system of regulation then is seen as providing some leverage for people particularly when confronted by unexpected economic loss. The argument is that legislators have been steadily replacing markets by courts in response to the wishes of voters for fair procedure in resource allocation decisions, and that this explains the recent history of regulation.

There appear to be two presumptions in this theory. First, that market risk cannot be avoided and second that there is little if any risk in being regulated. Both of these ideas seem dubious. It is exactly agents' risk aversion which gives rise to market institutions such as private insurance and joint-stock companies. These are designed to share and limit an individual's risk in any market and occur without the necessity for regulation. Alternatively, even though the process of regulation may be slower than market forces, there is no guarantee that the outcome will be any more or less favourable for a given individual than that achieved through the market. Regulation then delays risk but does not necessarily reduce it.

Owen and Braeutigam's theory seems to be based on the assumption that individuals would like to be able to avoid responsibility for their actions. Thus an individual wishes to take credit for any success but would like to be compensated for an unfavourable market outcome. While this is a believable inference about human nature, it is not a practicable way to govern an economy.

There are a variety of other theories which all stress the role of the private interest of particular groups in determining the form and substance of economic regulation.

(3) <u>Bureaucratic Preferences</u>

Niskanen [1968, 1971, 1975] seeks to explain the existence of regulation in terms of the preferences of the bureaucrats who staff the agencies. The behaviour of a regulatory agency, he suggests, can be interpreted as the administrators attempting to maximize their satisfaction in the face of constraints imposed by legal restrictions and opposing interest groups. Bureaux have two characteristics in this model. First, it is assumed that bureaucrats seek to maximize the total budget of the bureau. All arguments of a bureaucrat's utility function, such as salary, perquisites, reputation, power and prestige, are taken to be monotonically increasing in the size of the budget. Secondly, a bureau is assumed to exchange a specific output for a specific budget. This, it is argued,

> ... gives the bureau the same type of 'market' power as a monopoly that presents the market with an all or nothing choice. [Niskanen, 1968; 294]

This hypothesis would seem to provide an explanation for continual increases in regulatory budgets and largesse that is commonly perceived to exist in bureaucracies and quasi-autonomous non-government organisations (known as Quangos). It does not however, in this early form, adequately consider the demand side of the market for regulation. For instance, some bureaucrats are rewarded for making their bureau more efficient and for "trimming the fat" from the budget yet this is not compatible with the original theory. A later version [Niskanen, 1975] attempts to deal with this criticism. The utility function of the bureaucrat is slightly reformulated so that the objective is to maximize the "discretionary budget" which is the difference between total expenditure and the minimum cost of producing the expected amount. This gives the optimal quantity of regulatory services, as determined by bureaucratic preferences, which constitutes the supply of regulation. The demand side is assumed to come from vote-maximizing legislators.

The reformulated model is a model of bilateral monopoly in the market for regulation between the legislature and a bureau and, as such, the market equilibrium is indeterminate. Niskanen suggests that the process of regulation could be made more efficient by creating competitive behaviour on both sides of the market whereas Becker [1976, 1983] argues that it already exists.

Eckert [1973] echoes the Niskanen approach. However he contrasts the objectives of bureaucrats seeking to maximize their budget by extending the scope of regulation with those of the regulatory commissioners. Eckert suggests that regulatory commissioners have a commonality of interest with their respective regulated industries as they are frequently employed in the regulated industry upon the termination of their term with the regulatory body. This, he argues, is consistent with the creation and maintenance of regulated monopolies in otherwise competitive industries.

(4) Capture Theory

The main economic theory of regulation opposed to the naive public interest approach is the capture theory, of which it could be argued that Eckert [1973] is a variant. This agency-capture approach was developed and presented in various forms by Friedman [1962; 137-160], Stigler [1971] and Posner [1974]. The theory has since been extended and modified among others by Peltzman [1976] and Becker [1983]. The basic hypothesis of the theory is as follows.

> Regulation may be actively sought by an industry, or it may be thrust upon it ... as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit. [Stigler, 1971; 3]

The central idea of the theory is that the state has the power to coerce, which can be used to confer benefits or costs upon individuals and groups. Regulation is a commodity allocated by the laws of supply and demand. On the demand side, industry in particular is identified as being able to use the power of the state to increase its profitability through such measures as subsidies, entry control, price fixing and policies which affect substitute and complementary commodities. The other major group on the demand side of the market is that of consumers. Consumers' objectives, and hence demand patterns, depending on the form of ownership of the industries [Manning, 1986], may be expected to differ from those of industry. The state is the supplier of regulation and acts as an arbiter between the interest groups in determining the 'market equilibrium'. This role of the state within the political process is considered further when Peltzman [1976] and Becker [1976, 1983] are discussed later in the section.

The political decision process is different from the market in two important respects. First, because under a disaggregated system large numbers of agents must simultaneously make decisions, a process which would incur prohibitive transactions costs, decision making in the political system is commonly delegated to representatives. Secondly, the opportunity to absent from the effects of any decision is not readily available to an individual agent.

Conversely, those not directly involved have a potential voice in the decision. These factors combined imply that the political decision making process doesn't allow for participation in proportion to interest or knowledge, and, because of the existence of externalities, neither does it provide correct incentives for the acquisition of information on which to better base decisions [Stigler, 1971; 10-11].

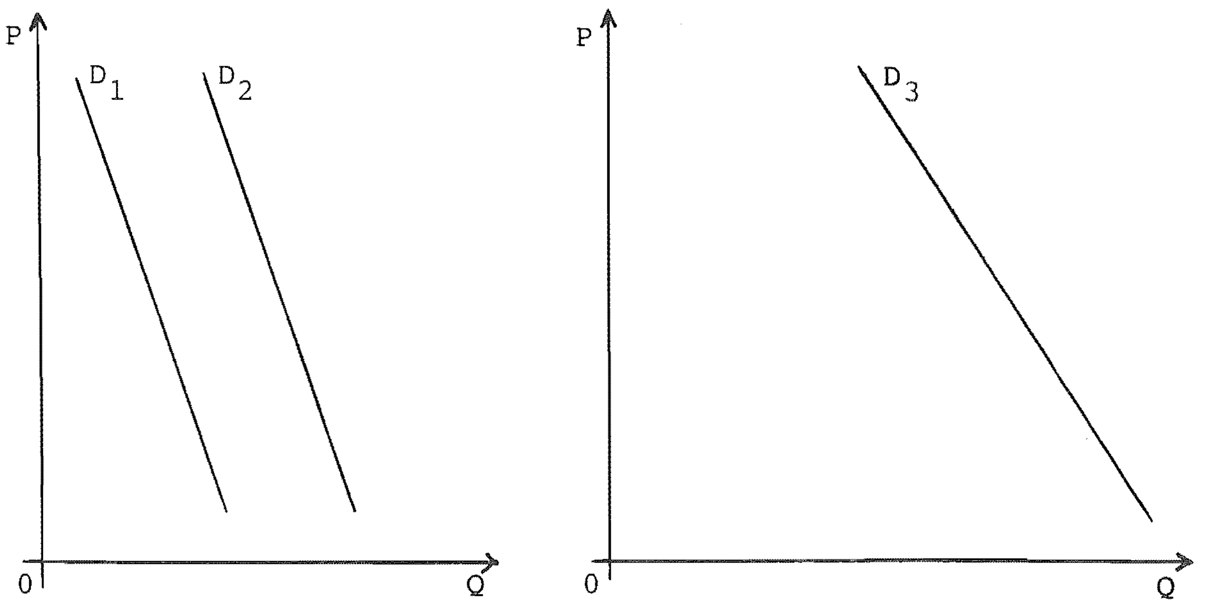
This frequently leads to situations with concentrated interest groups on one side of an argument and dispersed interests on the other. As industry interests are usually more concentrated than those of consumers, the consequence is that industry interests prevail often through intensive lobbying efforts and the 'buying' of legislation with inducements such as campaign contributions. This argument is illustrated in Figure 1-2-1 below.[8]

The interpretation of Figure 1-2-1 is as follows. Curves $D_1$ and $D_2$ in panel (i) show respectively the true and revealed demands of consumers for the regulation. Because of the dispersed nature of consumers, their interests are not communicated effectively to the regulation "market". Curve $D_3$ in panel (ii) shows industry demand. Summing horizontally in panel (iii), if $D_1$ were correctly revealed, the composite demand curve would be $D_4$ resulting in an equilibrium at Q*, the socially optimal quantity of regulation. Given however that only $D_2$ is revealed by consumers, curve $D_5$ emerges and the equilibrium regulation produced at $\hat{Q}$ is socially excessive.

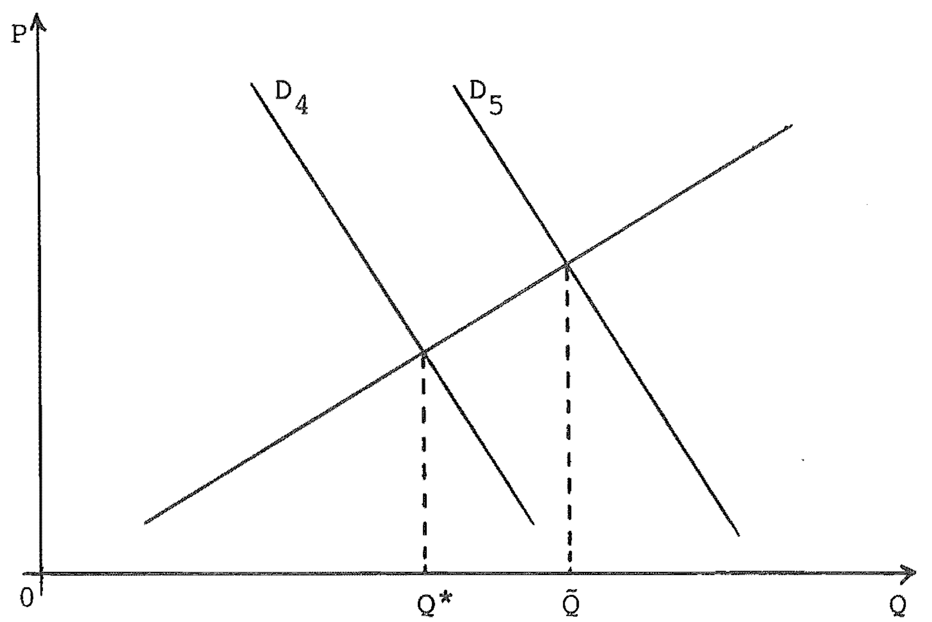
Under the extreme version of the capture theory, that which is implied by the earlier quote from Stigler, industry

Figure 1-2-1: Comparison of the equilibrium 'quantity' of regulation under actual and revealed demands

(i) Consumers' demand (ii) Producers' demand




(iii) Market equilibrium



interests always prevail. Regulation under this process may be viewed as somewhat akin to cartelization. A cartel attempts to increase the profits of its members by restricting output, the success of which depends on the elasticity of product demand and ease of entry to the industry. Against this benefit must be weighed the costs of cartelization which include the initial transactions costs of forming the agreement and the maintenance cost of the internal enforcement necessary to overcome the individual firm's incentive to exceed its quota allocation. Regulation uses the power of the state to accomplish the same objective thus externalizing some of these costs that are otherwise internal to the industry.

Posner [1974] suggests that the theory of cartels may provide some information about the supply and demand for regulation. He maintains however that, while cartels are relevant, they cannot form the basis of a theory of regulation because of a lack of historical empirical coincidence between the formation of cartels and the imposition of regulations. In particular he notes that the demand for regulation is often greater among industries for which private cartelization is prohibitively costly.

This observation is entirely consistent with the hypothesis of regulation as cartelization [Posner, 1974; 344-347]. It is likely that for some industries where large numbers of contractual parties are involved, the benefits of cartelization can be obtained much more cheaply through political lobbying for regulations, and that this option may make cartelization viable in situations where an industry-funded operation would be infeasible. Cost savings to the industry could occur at the initial bargaining or contract stage and with respect to the internal enforcement expenditure incurred during the operation of the cartel.

It is evident, however, that private cartel arrangements exist. This can be understood if the strict form of the capture theory is relaxed. Under the assumption that industry interests are fully reflected in the actions of the state, externally funded cartelization through regulation would be preferred to a private industry arrangement as the same benefit could be achieved at lower cost to the industry. If however it is assumed that regulation results from a compromise resolution of conflicting interests achieved through the political process, then cartelization by regulation confers benefits that are likely to be distributed differently and perhaps reduced in aggregate when compared with those obtained in a private profit-maximizing strategy. In this case, the costs to the industry of cartelization achieved through regulation are less than those of engaging in a private industry contract but so are the benefits. The pattern of regulation and private cartelization should therefore reflect the relative profitability of industries under the respective arrangements. If the theory is relevant then, it should be possible to explain the pattern of regulation and private cartelization using several key indicators of each industry such as the number of contractural parties.

Peltzman [1976] extends and formalizes the initial framework of Stigler and Posner. He argues that the capture theory is ultimately a theory of the optimum size of effective political coalitions set within a general model of the political process. His interpretation of the Stiglerian framework is that it views regulation as a political auction in which the highest bidder receives the right to tax the wealth of everyone else. Peltzman's model then seeks to explain why the successful bidder is usually a numerically compact group as in the case of competing producer and consumer interests. The key factor which emerges is the diminishing returns to group size which exist with respect to the costs of using the political process.

The basis of Peltzman's formulation is that the regulator wishes to maximize votes or more correctly, the majority margin.[9] In doing this, political candidates must choose the size of the beneficiary group, the amount of contributions required to mitigate opposition, and the size of the transfer to the beneficiaries. Among the first order conditions resulting from this framework is that the marginal political return from a transfer must equal the marginal political cost of the associated tax. Analysis of this condition reveals that the benefits to a group from government regulation are less than if a private cartel was formed because of the need to consider and placate opposing factions, a result which accords with the analysis of Stigler [1971; 6-7].

The conclusions that Peltzman draws from his formulation are twofold. First, the effect of imperfect information regarding the gains and losses of regulatory decisions, and the cost of political lobbying, is to limit the size of the winning group. This provides the justification for the claim that compact groupings, such as industry interests, will prevail at the expense of dispersed interests. Second, as argued above, the gain to the winning group is not as great as the political process could grant because political entrepreneurship will produce a compromise or coalition of interest groups which includes members of the losing group or groups. That is, the winning group is unable to capture all of the gains from trade through the political process. The fact that rational maximizing agents pursue regulation however suggests, as earlier implied, that the reduction in benefits granted is not as large as the cost saving which the state organization provides.

Becker [1976; 245-246] shows that Peltzman's model requires the existence of a positive deadweight loss. Concentrating on this aspect of the model and assuming that "voters perceive correctly the gains and losses from all policies" [Becker, 1976; 247], he argues that regulatory policies such as quotas are used because they involve smaller deadweight losses than cash transfers and suggests that

> ... industries are regulated ... because industrial regulation may be a relatively efficient way of transferring benefits to specified groups. [Becker, 1976; 247]

Becker [1983] extends this analysis in a model of competition among pressure groups. The central results of this model are that the successful groups are small relative to the size of the groups that are taxed to pay the subsidies [Becker, 1983; 385] and that competition amongst the groups generates controls that are efficient methods of conferring benefits on the winning group [Becker, 1983; 386]. These results emerge from the effects of marginal changes in deadweight loss on the political pressure exerted by the competing groups.

A final interesting variant of the producer-interest theory of regulation, that illustrates the point about compact groups winning through the regulatory process, is that of Salop and Schefman [1983]. Their analysis considers the case of a dominant firm, facing a competitive fringe, which uses regulation to raise its rivals' costs.

> Disadvantaging competitors can provide a benefit that exceeds its costs, if the strategy allows the dominant firm to increase price or market share. [Salop and Schefman, 1983; 268]

They examine the effect of regulations on the industry demand curve, the fringe supply curve and the residual demand and average cost curves for the dominant firm. The dominant firm can use the political arena to inflict costly regulation on its rivals and possibly even on itself. A sufficient condition for this to be a profitable strategy is if it shifts the dominant firm's residual demand curve up by more than its average cost curve at the original output level. This shift depends on the product elasticity of demand facing the industry and the elasticity of the fringe supply curve. Empirical justification for this hypothesis is alluded to in Section 1-3.

(5) Life Cycle

The final approach considered here is the Life-Cycle hypothesis [Bernstein, 1977] which concentrates on the evolution of regulatory behaviour over time. Bernstein [1977; 11-12] argues that short term coalitions of consumer interest groups form to have regulations passed in the legislature through political pressure. As time goes by, the issue loses its political importance partly because of a sense of accomplishment, the coalition fades in potency, and the regulatory machinery becomes captured by the industry to use for its own purposes.

One criticism of this hypothesis is its generality. Too many different regulatory patterns can fit tolerably well with this explanation, but no particular behaviour is predictable. This means that there is no basis for testability and empirical verification. It is also strange that the possibility of a once powerful consumer action group reforming when it observes outcomes contrary to its objectives, is ignored in this framework. This is in contrast to the competitive nature of the political process alluded to by Stigler and Peltzman and extended by Becker. Under Becker's assumptions such a change in regulation could result only from a fundamental change in the perceived deadweight losses of regulation to the respective interest groups. This is because the change in regulatory behaviour would cause a significant redistribution of wealth amongst the competing interest groups and

> ... even heavily taxed groups can raise their influence and cut their taxes by additional expenditures on political activities. [Becker, 1983; 372]

## 1-3 COSTS AND EFFECTS OF ECONOMIC REGULATION

Another branch of the regulation literature is concerned not so much with the explanation of the existence of regulations, but rather with its effect on economic agents, particularly firms and individuals directly subject to it.

One of the most significant contributions in this area was made by Averch and Johnson [1962]. Their article on the behaviour of a monopoly under a rate of return regulatory constraint spawned a proliferation of research and comment on the so named Averch-Johnson effect commonly referred to in the literature as the A-J effect.[10]

This A-J effect is concerned with capital padding under certain conditions. Their article shows that if the allowed rate of return is greater than the cost of capital, but less than the unconstrained profit maximizing rate, then the firm will substitute capital for the other factor of production, which is usually labour in a two factor model, and operate with factor proportions that are not cost minimizing for that level of output. Operating in this way is not efficient from an economy wide, resource use viewpoint, but it is a rational response for the firm within the regulatory context as it minimizes reductions in profit.

The importance of the work of Averch and Johnson cannot be overstated; not so much for the result derived but because, by demonstrating that under certain assumptions regulation could induce unintended detrimental side effects, it provided a focal point for intellectual disenchantment with regulation. Given the knowledge that these potential side effects existed, the question of the cost effectiveness of regulations arose. This prompted studies that attempted to identify and quantify the effects of particular policies in order to derive some measure of the costs of specific regulations.

Most studies use an incremental cost approach where the costs attributed to the regulation are measured as the difference between the costs that occur in the presence of regulation and those that would exist in its absence. Deciding on the base level involves a certain degree of value judgement. In the area of compliance costs for example, not all expenditure which seems related to the regulation is necessarily able to be correctly attributed to it. It is possible that the agent would have engaged in some voluntary expenditure in the same area even in the absence of any compulsion to do so [Kosters, 1979; 18-19].

Several types of costs are commonly identified in these studies including administrative costs, compliance costs, and induced or indirect effects. Administrative costs are those costs covered by the budgets of the regulatory agencies, which are incurred by government and financed out of general taxation. Compliance costs are those costs incurred by firms subject to the regulations which are directly attributable to the regulations. These include such costs as designing, implementing, and operating changes in productive processes to meet new standards, and also the costs involved in maintaining records of performance that may be required to demonstrate compliance.

In one such study, Simon [1981] found that in 1977, for the corporations covered by the study, the overall effect of federal regulations was equivalent to an increase in the corporate tax rate of approximately fourteen percent, and also that the effects were very disparate between different types of industries. These increased costs are ultimately passed on to consumers in the form of higher prices or, where necessary, to the shareholder in the form of reduced dividend payments, and represent a "hidden tax" of the regulation that is transferred from the government to individual agents in the economy [Weidenbaum, 1980; 6].

It may appear as if such cost increases are contrary to the predictions of the various theories of regulation. A regulation-induced price increase for instance reduces consumer surplus and thus may be thought to be against the public interest. If however the commodity in question generates a negative externality, the price increase, by reducing consumption and hence, in equilibrium, production of the good, is quite possibly in the public interest. Alternatively a policy which increases industry costs appears to be inconsistent with the results of the capture theory, at least in its strict form.[11] Maloney and McCormick [1982; 104] however identify conditions under which regulation-induced cost increases enhance industry profitability and, as stated earlier, Salop and Schefman [1983] show that it is possible for cost inflation to benefit the dominant subset of an industry.

The third major category of costs are those which arise from the indirect or induced effects of regulation. These costs are the source of most of the deadweight losses attributable to regulation because they lead to inefficiency in production and consumption. As Weidenbaum notes;

> While costs arising from inefficiency are more subtle to grasp and more difficult to measure than the other types of costs, they are in a sense the most serious because they represent pure waste of resources. [Weidenbaum, 1980; 24]

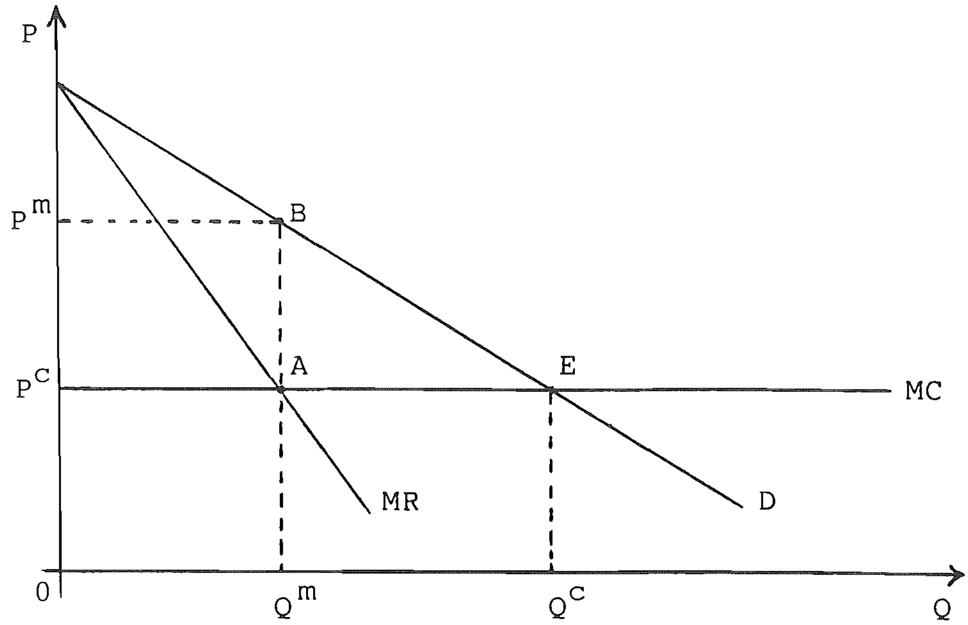
These effects include firstly, less product research and development, and a reduced rate of introduction of new products. Under some regulations, the length and costs of the approval process make the creation of new products less viable. According to Peltzman [1977] this is especially true in the pharmaceutical industry where Food and Drug Administration controls in the United States delay or prevent the introduction of many new drugs. Secondly, in many industries, the compliance costs of meeting government requirements are so great that expenditures in this area replace productive investment in new plant and equipment. Thirdly, in some instances, regulations directly distort the market in which they apply. An example of this is minimum wage legislation which prices low productivity labour out of the market and necessitates further government involvement in the form of unemployment benefits.[12]

Fourthly, activities such as lobbying, which are undertaken in an effort to procure regulation, represent additional costs of the regulatory process. Such activities are usually categorized as rent-seeking behaviour. Tollison [1982] defines rent-seeking as "the expenditure of scarce resources to capture an artificially created transfer" [Tollison, 1982; 576].

The theory of rent-seeking assumes that the process of acquiring the artificially created rents is competitive. An illustration commonly used in the literature[13] is that of monopolization in a competitive industry. This is shown in Figure 1-3-1 below.

The diagram shows an industry with constant marginal costs. Under competition equilibrium output is $Q^c$ and zero profit is generated in the industry. If monopolisation occurs, output is restricted to $Q^m$ where marginal revenue

Figure 1-3-1: The costs of monopoly




and marginal cost are equated. The usual analysis assumes that the rectangle $P^cP^mBA$ represents a simple transfer of surplus from consumers to producers and thus shows industry profits under monopoly. The triangular area ABE represents the deadweight loss of monopoly. This is the amount of lost consumer surplus that is not appropriated by producers.

The rent-seeking literature holds that expenditures incurred in the process of attaining monopolistic control of the industry aggravate the social cost of monopoly. Posner [1975] indeed argues that the entire trapezoidal area $P^cP^mBE$ represents the cost to society of monopolisation. Much debate exists in the literature between this type of approach which states that all such expenditures are directly unproductive, [see, for example, Bhagwati, 1982], and others which suggest that, by entering the utility functions of some agents, these expenditures do not represent a total loss to society in their entirety, [see for example

Brooks and Heijdra, 1986]. Regardless of the quantitative merits of these respective arguments, however, it is evident that, to the extent which the process of currying favours through the political system is competitive, the usual triangular deadweight-loss measure underestimates the social costs of obtaining monopoly through regulation.

Another possible effect, which is central to the analysis of Salop and Schefman [1983], is that the burden of a regulation may be disproportionately heavy on smaller enterprises because of diseconomies in paper work and other legalities of compliance. In effect, regulation may add a fixed cost to production. In this case, the cost structure of these small firms is biased upwards relative to larger entities. If this effect is significant enough, the small firms will become uncompetitive, and greater concentration in the industry concerned will result in increased monopolistic distortions.

Finally, there is what Weidenbaum labels the "bureaucratization of corporate activity" [Weidenbaum, 1980; 13], whereby management's attention is diverted, to a greater or lesser extent, from normal activities such as product development, production, and marketing, to implementing government mandated requirements. This change of focus strikes at the very heart of the private enterprise system, the entrepreneurial spirit, and, if the redirection is significant, the effects on patterns of production and consumption in the economy could be far reaching and severe albeit difficult to measure.

## 1-4 ENFORCEMENT AND COMPLIANCE

The regulatory costs and distortions presented above are extensive but by no means exhaustive. Most of the theories and empirical studies discussed in Sections 1-2 and 1-3 appear to implicitly assume that the regulatory constraint is binding, absolute and inviolate. The possibility of illegal evasion of the constraint is either ignored or quickly skirted. Indeed, Smith [1976] states when discussing other theoretical treatments of regulation that;

> ... analyses have in general neglected the conflict between the incentive for individuals to evade regulations and the effort of governments to enforce regulations. [Smith, 1976; 394]

a sentiment echoed by Veljanowski [1983].

> Enforcement is a neglected topic in the economics of regulation. The formal analysis of regulatory standards invariably assumes that all firms comply with the law and, implicitly, that enforcement is costless and complete. [Veljanowski, 1983; 123]

This assumption however largely flies in the face of reality. Almost without exception, enabling legislation for regulation contains provisions for the imposition of penalties.

By its very nature, a regulation is designed to modify existing behaviour or limit future courses of action. If this were not so, the measure is redundant and would not be enacted. Assuming however that economic agents are rational, individual decisions prior to the introduction of any outside control would be made on the basis of profit or utility maximization at the private level irrespective of any "market failure" which may or may not exist. Compliance with a regulatory constraint then involves a reduction in either profit or utility providing a clear incentive for evasion by some agents and thus "the law must be enforced if it is to have any impact" [Veljanowski, 1984; 171].

The incentive to evade is apparent no matter which theory of economic regulation is advanced. It is obvious that producers would seek to evade regulations of the pure public interest type enacted on behalf of consumer groups, but even under the strict interpretation of the capture theory there is an incentive to evade on the part of the individual firms within the industry, analogous to that which gives rise to the internal policing problem in a cartel.

In these circumstances, questions relating to the methods and costs of enforcement are clearly relevant to the regulation decision.

> Enforcement difficulties, increasingly significant, reduce the net benefits from government regulations, and should affect choices about whether and/or how to regulate. [McKean, 1980; 269]

Recently, discussions and theoretical analyses of enforcement have appeared in the regulation literature. In one such discussion McKean [1980] identifies factors which influence enforcement including the measurability of violations and the ability to identify their source, the number of agents subject to the regulation, the size of the penalty, and the elasticity of demand for the product being regulated. This last factor is significant in determining public support for regulations; itself a significant issue.

> The need for public support as a prerequisite for enforcing economic controls seems obvious. [Bernstein, 1977; 217]

> If the case for regulation is comparatively clear and a substantial majority of the public approves of it, those being regulated experience extra pressure to accept the rule instead of bargaining, and to obey the rule instead of violating it. As a consequence, compliance will be comparatively high. [McKean, 1980; 285]

Several other writers discuss the minutiae and practicalities of regulatory enforcement in a similar manner.[14] Valjanowski [1983] discusses what he refers to as the "compliance enforcement system" which consists of a process of direct negotiations and cooperative bargaining between regulator and regulatees that results in "discretionary flexible enforcement" [Veljanowski, 1983; 123]. Regulatory offenses result not so much from the calculated risks of a profit-maximizing firm as from the genuine mistakes of a bumbling organizational structure and regulatory agencies are not a police force whose function is detection and prosecution but rather engage in "bargaining in the shadow of the law" [Veljanowski, 1983; 126].

To this multi-dimensional model of Veljanowski, Kagan and Scholz [1984] add that non-compliance may be the result of principled disagreement by a concerned political citizen and argue that

> ... one implication of the diverse sources of non-compliance is that indiscriminate reliance on any single theory of noncompliance is likely to be wrong, and when translated into an enforcement strategy, is likely to be counterproductive. [Kagan and Scholz, 1984; 85]

These discussions are interesting and contain many relevant and important points. They cannot escape however the central idea that regulation must be enforced. Several articles concentrate on this in presenting more theoretical analyses of regulatory enforcement.[15] Such models almost exclusively concentrate on a partial equilibrium approach to the problem of regulation and enforcement. Much of the analytical framework used in these models stems from the pathbreaking article on the economics of crime by Becker [1968]. In Becker's approach, the decision to engage in crime, as with any other economic activity, is based on an evaluation of the costs and benefits involved. With outcomes uncertain, an individual engages in those activities which maximize expected utility.[16]

Most of the regulatory models derive that an agent's compliance with a regulatory constraint is determined where the marginal cost of noncompliance is equal to the marginal cost of so doing and that under an enforced regulatory constraint

> ... socially optimal control is found to occur at the point where marginal damages equal marginal control costs plus the marginal costs of enforcing these controls. [Linder and McBride , 1984; 328]

Lee [1984] argues in addition that

> ... by substituting harsh penalties for effective enforcement the deterrent effect of a given expected penalty can be maintained at less cost. [Lee, 1984; 153]

This reduces the marginal cost of enforcement and allows an increase in the extent of socially optimal control. The logical conclusion is then to have an infinite fine with no enforcement. In this situation enforcement costs are zero and hence control and compliance can be complete.[17] In practice,however, fines are not infinite because of the necessity, as Stigler [1970] notes, to maintain marginal deterrents sufficient to prevent "spillovers" between different classes of offense. Enforcement therefore requires resources and thus

> There is one decisive reason why the society must forego "complete" enforcement of the rule: enforcement is costly. [Stigler, 1970; 124]

## 1-5 AIMS AND SYNOPSIS

Discussion of "optimal" behaviour raises questions. Optimal behaviour typically results from the constrained optimization of some objective function. Relevant questions therefore concern the arguments of the objective function and the optimization strategy.

Polinsky [1979] argues that enforcement agents' objectives are most likely to be different from those which constitute social welfare while Veljanowski [1984] states that

> ... the observation that regulatory law is often not vigourously enforced indicates to proponents of a penalty system government insincerity and agency capture, if not corruption. [Veljanowski, 1984; 176]

and Diver [1980] notes that enforcement agencies "concentrate on the relatively most trivial offenses to the exclusion of the most serious" [Diver, 1980; 260].

No attempt is made in the literature to reconcile these differing observations concerning penalty structures and enforcement practice in any consistent theoretical model. Perhaps this is because there is no consistent basis to compare the various theories of regulation either.

Although the models outlined in Section 1-2 have become increasingly sophisticated, none, of itself, is capable of explaining observed variations in regulations, and there are no clear criteria which can be used to categorize a particular regulation as belonging to one approach or another. It may be possible in certain cases to infer regulatory objectives from a reinterpretation of historical evidence, [Kolko, 1863; 98-108; Rose, 1985], but this luxury is seldom afforded when dealing with contemporary problems.

To make a valid comparison of the various approaches to explaining regulation each must be embedded in the same theoretical structure and the results generated by the models evaluated and contrasted. This is the first aim of the present thesis. A simple structural model of an enforced regulation is developed, initially within a partial equilibrium framework and later in a general equilibrium model of a simple economy. Into this otherwise identical behavioural model are embedded two contrasting regulatory hypotheses; the Naive Public Interest Theory (NPIT) and the Capture Theory (CT). The behavioural implications resulting from the hypotheses can then be derived and compared.

The second aim of the thesis concerns these behavioural implications of the varying regulatory hypotheses. Each hypothesis contains the objectives of the regulator. On the basis of these objectives the behavioural implications which emerge from the optimization process of the model may be termed to be optimal for that regulator. This emergent behaviour includes the penalty structure and enforcement practice. Working in reverse it is then possible to infer regulatory objectives from observed behaviour with respect to the decision variables of the regulator.[18] A theoretical structure is established therefore whereby it is possible to consistently explain variations in enforcement practice and to infer objectives from observed enforcement behaviour. This is the second aim of the thesis.

The great majority of analyses of regulation, and particularly of those which explicitly consider enforcement, are set in a partial equilibrium framework. Regulation and its enforcement are likely, however, to affect sectors of an economy other than that which is directly controlled.

These impacts in other sectors may then feed back into the regulated sector and alter the effectiveness of the regulation in the sector to which it directly applies. The implications for regulatory policy of such a process cannot be explored within a partial equilibrium approach. Because of this, the thesis includes the construction of a simple two-sector general equilibrium model. This model is used to examine whether the results derived in the partial equilibrium framework hold true in a general equilibrium context. The structure of the thesis is as follows.

Chapter Two develops the partial model of regulation which consists of a tax or quota used to control a negative externality generated by a competitive industry. Given that the regulatory instrument is enforced by an expected monetary penalty, a regulated equilibrium is defined and its properties examined.

Chapter Three incorporates the two competing hypotheses concerning regulatory objectives and derives optimal regulatory policies in each case. The responses of these policies to parameter changes are explored and the results compared and contrasted.

Chapter Four provides an application of this partial equilibrium model to another area of market failure, namely the common-property externality as portrayed through the example of an open-access fishery. A brief literature review precedes the development of the fisheries model and the application of the regulatory constraint to it. Here the assumption is that regulation is by means of enforced output quotas. The behavioural implications of the two hypotheses are compared and contrasted.

Chapter Five incorporates the regulatory framework within a simple two sector general equilibrium model of a Ricardian economy. As previously a regulated equilibrium is defined but in addition the economy's locus of all regulated equilibria is derived. The inclusion of the regulatory hypotheses determines the optimal regulatory policies which correspond to differing points on the locus of regulated equilibria. The results are presented in general form and also using a Cobb-Douglas utility function for illustrative purposes.

Chapter Six concludes and indicates areas for future development.

NOTES

1. Subsequently, notable theorists, including Milton Friedman, have argued that state intervention with the money supply was responsible for the Depression.

> From the cyclical peak in August 1929 to the cyclical trough in March 1933, the stock of money fell by over a third ... The monetary collapse was not the inescapable consequence of other forces but rather a largely independent factor which exerted a powerful influence on the course of events ... it is hardly conceivable that money income could have declined by over one-half and prices by over one-third in the course of four years if there had been no decline in the stock of money. [Friedman and Schwartz, 1963; 299-301]

2. See for example Averch and Johnson, 1962; Friedman, 1962; Posner, 1969 and 1974; Stigler, 1971; and Stigler and Friedland, 1962. These and other approaches to regulation are outlined in the following sections of the chapter.

3. This approach is similar in spirit to the "Life-cycle" hypothesis of Bernstein [1955] which is presented later in the section.

4. See for example Peltzman, 1976; and Becker, 1983.

5. The authors' acknowledge that "justice" and "fairness" are "fuzzy" concepts. They define these terms as "treating equals equally" and emphasize "procedural fairness" as the ability to "rank economically identical outcomes on the basis of the manner in which they were attained" [Owen and Braeutigam, 1978; 20].

6. An alternative hypothesis is of course that the ideological principle is not widely shared.

7. An illustration of this in New Zealand occurs in the area of environmental regulation. In this case

development proposals which will adversely affect the environment must go through the Planning Tribunal where concessions are often granted to conservationists and minority ethnic groups who, of themselves, have little economic power.

8. The demand curves are drawn as shown for illustrative purposes only. Precise conditions on their shape, slope, and position, would require careful consideration which is not attempted here. The horizontal axis shows some index of the degree of intervention. The diagram reflects the assumption that the regulation favours producers at the expense of consumers.

9. Assuming that the regulator and politician are identically one, Hirschliefer [1976] argues that the utility function of the regulator is more than uni-dimensional. In particular he suggests that wealth maximization is a more appropriate objective. Given this, the majority margin is only one argument of the objective function. A politician or regulator can be induced by an appropriate amount of money to accept a greater risk of defeat; that is a smaller majority. There are diminishing returns then to the size of the majority.

A multivariate objective function requires modifications to the first-order conditions present in the Chapter. It does not, however, alter the fundamental result that concentrated interests prevail in the political process at the expense of more diffuse interest groups as typically the concentrated interest groups are those able to make some form of "compensatory payment" for the risk that the politician incurs in supporting their cause.

10. Peltzman [1976; 224] notes that the regulation of both natural monopolies and naturally competitive industries can be consistent with regulation delivered through the political process. The prevailing interest group in each case may be expected to differ.

11. That is, using the assumption that the industry captures all the benefits of regulation.

12. In New Zealand rigidities in the labour market are caused not so much by government-set minimum wage levels but by union-negotiated award wages. Unions are structured on a craft basis and often have jurisdiction over a number of industries as membership is compulsory. The award wage sets a floor on the wage that must be paid to members of the union and thus creates distortions in the labour market both within and between industries which employ unionised labour.

13. See for example Brooks and Heijdra, 1986; Posner, 1975; Putsay, 1978; Tollison, 1982; and Tullock, 1967.

14. See for example Diver, 1980; Hawkins and Thomas, 1984; Kagan and Scholz, 1984; and Veljanowski, 1983 and 1984.

15. These include Harford, 1978; Lee, 1984; Linder and McBride, 1984; Papps, 1985; Smith, 1976; Storey and McCabe, 1980; and Viscusi and Zeckhauser, 1979.

16. Subsequent extensions and developments of Becker's model have been presented by Ehrlich, 1972, 1973 and 1982 ; Kemp and Ng, 1979; Polinsky and Shavell, 1979; and Stigler, 1970; among others.

17. The result depends on defining the limit of the expected penalty which is the product of an infinite fine multiplied by a zero probability of detection. From limit theory, this product can take any value depending on the rates of convergence of its component terms. The implications of this "optimal penalty" result are discussed further in Chapter Two, Section 2-7. Suffice it to say here that infinite fines are not observed in practice.

18. Ross [1984, 1985] poses a similar question in a model without enforcement. Ross assumes that the regulator seeks to maximize some weighted sum of consumers' and producers' surpluses, subject to a profit constraint for the regulated form. The weights are not directly observable but can be imputed by using data on prices, marginal costs, and demand elasticities.

# CHAPTER TWO

## REGULATION AND ENFORCEMENT: A PARTIAL EQUILIBRIUM APPROACH

### 2-1 INTRODUCTION

In Chapter One Section 1-1 it was observed that the necessary conditions required for the perfectly competitive paradigm are seldom, if ever, satisfied in real economies. The existence of factors such as externalities, increasing returns, market concentration, and public goods, causes the private and social costs of economic activity to diverge and invalidates the efficiency results of the first-best market outcome.

In these circumstances, governments in mixed economies often seek to either indirectly supplement or directly alter market outcomes via any of a plethora of policy instruments including taxes, subsidies, price and profit controls, entry restrictions, and output quotas. The relevant questions for policymakers and interest groups concern firstly which, if any, of these instruments to use and secondly the appropriate degree of control to exercise. The answers depend on the objectives of the regulator[1] and an evaluation of the relative performance of intervention, as compared with the original market outcome, in achieving them.

The problem of whether and how to control a particular market then requires an understanding of the way in which the introduction of a regulatory constraint affects agents' decisions within the market. It is this question which the present chapter addresses.

The chapter begins by characterizing the competitive market outcome in a partial equilibrium framework showing the incentives of individual agents within the market and how they interact to produce the aggregate result. The competitive equilibrium is assumed to diverge from the socially optimal or first-best outcome and regulatory instruments are introduced to correct this imbalance.

Discussion has occurred in the literature on the relevant merits of 'price-oriented' instruments, such as taxes and subsidies, and quantitative restrictions such as licencing and quotas.[2] In light of this, the present analysis employs one instrument of each type; a unit sales tax and an output quota. The effects of these policies on individual decisions and the market outcome are derived after introducing the concept of a regulated equilibrium.

Some of the many differing theories of regulation that exist in the literature were outlined in Chapter One, Section 1-2. Of these, two are used in this analysis; the Naive Public Interest Theory (NPIT) and Capture Theory (CT). The regulated equilibrium structure provides a consistent basis for the incorporation of the differing hypotheses concerning regulatory objectives which arise from the two theories, and allows for a valid comparison of their predicted outcomes.

## 2-2 DERIVATION OF THE UNREGULATED COMPETITIVE EQUILIBRIUM

The economy is assumed to comprise many different industries each producing one unique commodity. One industry visits a negative externality on other industries in the economy. This industry is analysed here. In the present partial equilibrium context, the rest of the economy is ignored apart from consideration of the effects of the externality.

The demand side of the market for the commodity produced in the externality-generating industry is assumed to consist of a finite number of individual consumers with identical well-ordered preferences. Individual demands are assumed to be decreasing in price and differentiable.

$$(2\text{-}2\text{-}1) \qquad q_i^d = d_i(P, y_i) \; ; \; \frac{\partial d_i}{\partial P} < 0 \; , \qquad i=1,\ldots n$$

represents the quantity demanded of the commodity by the $i^{th}$ consumer where P is the price of the commodity and $y_i$ the consumer's income.

Summation over all individual demand curves gives the market demand curve for the commodity which is itself continuous and negatively sloped. It is assumed that individual demand functions exhibit identical income effects, therefore

$$(2\text{-}2\text{-}2) \qquad Q^d = d_1(P, y_1) + d_2(P, y_2) + \ldots + d_n(P, y_n)$$

$$= D(P, Y) \; ; \; \frac{\partial D}{\partial P} < 0$$

where $Y = \sum_{i=1}^{n} y_i$ is aggregate income.

The existence of a single market equilibrium price together with the negatively sloped market demand curve causes a divergence between price and value in consumption on all intra-marginal units of the commodity known as Marshallian consumer surplus. This measure of aggregate surplus will be used in assessing the welfare effect of regulatory actions.[3]

The area under the market demand curve at any quantity represents the aggregate valuation of that quantity of the commodity by all consumers. The market demand curve therefore shows the marginal valuation of the commodity in consumption and may be termed the marginal social benefit (MSB) curve.

It is assumed that the supply side of the market comprises a finite number of individual firms with identical increasing marginal cost curves. An individual producer's supply curve shows the maximum quantity that the producer will willingly supply at various prices or alternatively, the minimum price that the producer must receive in order to induce the production of a particular unit of output. This is derived on the basis of profit maximization which involves maximizing revenue subject to the constraints of technology and factor prices embodied in the cost function.

For the representative firm in the industry facing a fixed market price, profit is given by the following expression.

$$\pi_i(q_i) = P.q_i - c(q_i) \qquad i=1,\ldots m \tag{2-2-3}$$

where $q_i$ represents the amount of the commodity produced by the $i^{th}$ firm in the industry.

Differentiating (2-2-3) with respect to output and assuming an interior solution gives the first-order condition for profit maximization

$$\frac{\partial \pi_i}{\partial q_i} = P - c'(q_i) = 0 \tag{2-2-4}$$

Profit maximization requires that, in the neighbourhood of the optimal $q_i$, $c''(q_i) > 0$:, the firm's cost function is convex. This in turn implies that the profit function is strictly concave and ensures that the solution to (2-2-4) is unique.

Assuming that the profit function is globally concave, the solutions to (2-2-4) over all P map out the individual firm's supply curve

$$q_i^s = s_i(P) \quad , \qquad i=1,\ldots m \tag{2-2-5}$$

where $q_i^s$ represents the quantity supplied of the commodity by the $i^{th}$ firm in the industry.

With strictly increasing marginal costs, individual supply is increasing in price. Totally differentiating (2-2-4) and rearranging gives

$$\frac{dq_i^s}{dP} = \frac{1}{c''(q_i)} > 0 \quad , \qquad i=1,\ldots m \tag{2-2-6}$$

Using the assumptions here, the individual firm's supply curve is its entire marginal cost curve. The properties of the industry supply curve depend on assumptions made regarding the size of individual firms in relation to the total market, and entry conditions. With free entry, industry and firm profits will be either zero or arbitrarily close to it. If the market equilibrium quantity is a multiple of the minimum efficient individual output level, equilibrium price will be equal to the individual firm's minimum average cost and profits, individual and aggregate, will be zero. If this is not the case, the price will be such that the marginal firm cannot profitably enter the industry but all incumbent firms make arbitrarily small positive profits. The long run industry supply curve is then essentially horizontal at a price in the neighbourhood of the minimum average cost level.

If free entry does not exist, any expansion in industry output requires expansion by individual firms which, from (2-2-6), requires an increase in market price.

It is assumed in the following analysis that there are sufficient firms for the industry to be competitive in that each firm acts as if it faces a fixed price in making its decisions but not perfectly competitive in the sense that entry is not free. In this case therefore the market supply curve is positively sloped thus

$$(2\text{-}2\text{-}7) \qquad Q^S = s_1(P) + s_2(P) + \ldots s_m(P)$$
$$= S(P) \quad ; \quad \frac{dS}{dP} > 0$$

The upward sloping nature of the market supply curve together with a single market equilibrium price implies that, at any positive equilibrium quantity, the price exceeds the supply reservation price on intramarginal units and thus positive economic profits or producer surplus is generated in the industry.

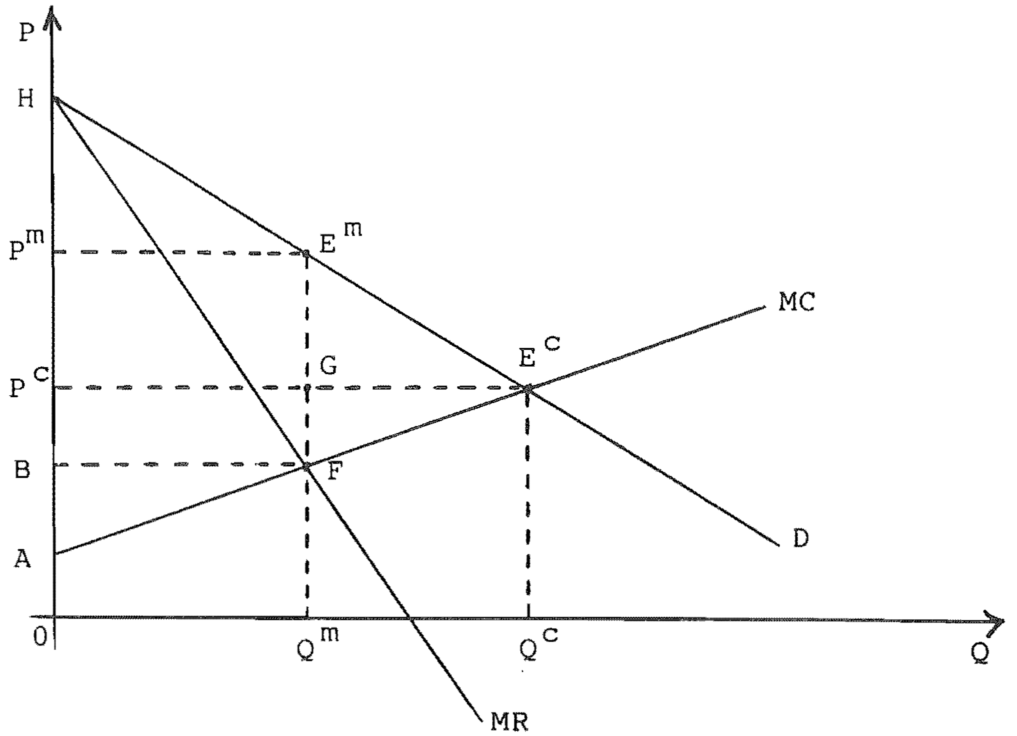
Market equilibrium in an unregulated environment occurs at the price which equates aggregate supply and demand decisions. Thus

$$(2\text{-}2\text{-}8) \qquad S(P^*) = D(P^*)$$

where $P^*$ is the equilibrium price of the commodity.

At this price $P^*$ all consumers are maximizing utility subject to their budget constraint and all producers are maximizing profits subject to the technological and financial constraints on production. In unregulated competitive equilibrium all individual plans are mutually consistent and the only constraints upon behaviour are those of the market. Figure 2-2-1 below illustrates industry equilibrium in the unregulated environment.[4]

Figure 2-2-1: Unregulated competitive equilibrium

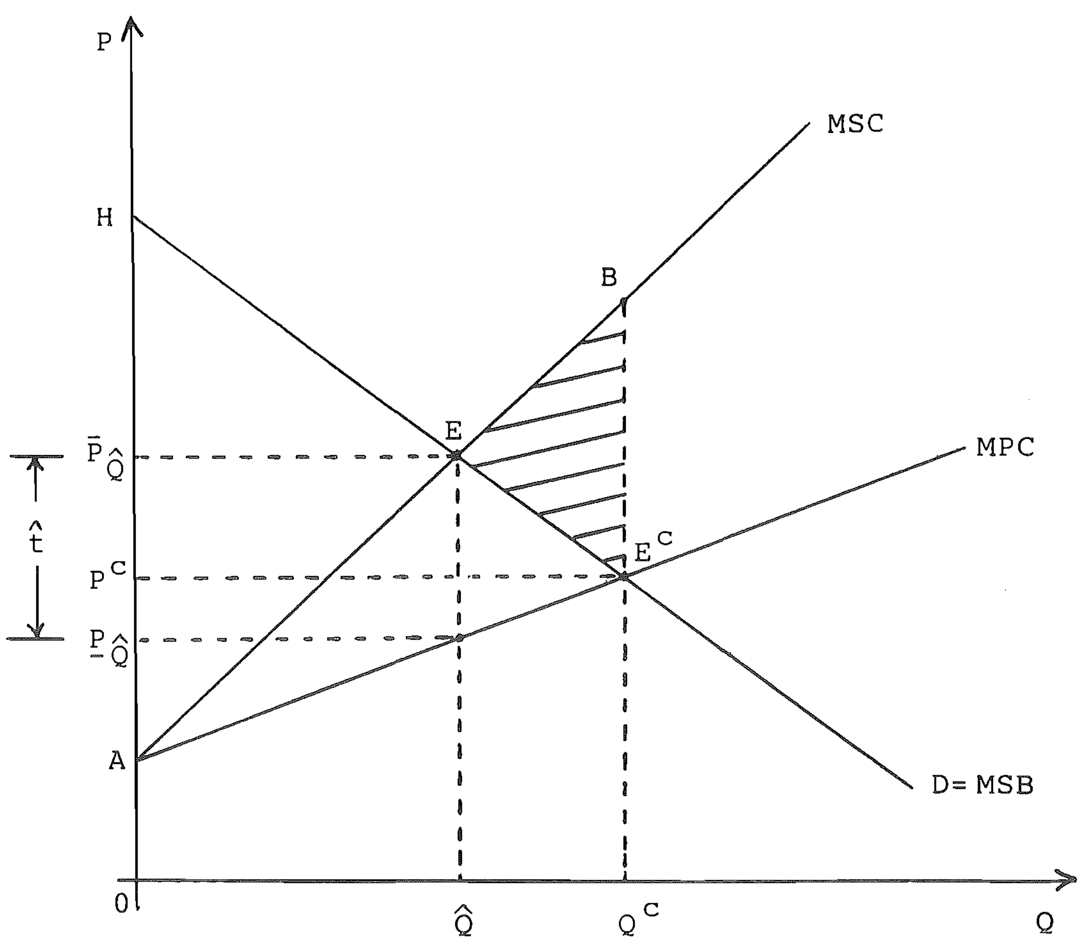



The competitive equilibrium occurs at $E^c$ where price equals marginal cost of production and aggregate surplus is maximized at magnitude $AE^cH$. No individual firm has an incentive to restrict output as this would merely serve to reduce its profit. Under a cartel or legalized monopoly however, output is restricted to the level $Q^m$ where marginal revenue equals marginal cost. At this output level, the industry enjoys profits of $P^mE^mFA$ which exceed the competitive level of $P^cE^cA$. The demand price or marginal social benefit of the commodity exceeds the marginal cost. Assuming that the marginal cost curve in Figure 2-2-1 reflects true marginal social cost, monopolistic control of the industry results in an aggregate welfare loss of area $FE^mE^c$ when compared with the competitive outcome.

Frequently, however, for reasons such as those outlined in Section 2-1, it is not the case that marginal private costs and marginal social costs of production coincide. The implications of the assumption that the output produced in the industry exerts a negative externality on the rest of the economy are now examined.

Increased output in this industry reduces the productivity of factors employed elsewhere. This represents a cost to the economy as a whole which is not reflected in the private decisions in this market. The marginal private cost (MPC) understates the true marginal social cost (MSC) of production as shown in Figure 2-2-2 below.

Figure 2-2-2: Unregulated competitive equilibrium with an externality



The competitive equilibrium $E^C$ involves a greater industry output than the socially optimal level $\hat{Q}$ where the marginal social benefit and marginal social cost are equal. The competitive outcome therefore produces an aggregate welfare loss of area $\hat{E}BE^C$.

Figure 2-2-2 illustrates the competitive equilibrium in the industry which generates the negative externality. It is possible to envisage the case when monopoly control would also result in an output greater than $\hat{Q}$. It is true, however, with constant or decreasing returns to scale in the production of a commodity which has a decreasing marginal social benefit and non-decreasing marginal social cost, that restriction of output by a monopolist reduces the size of the social loss from that associated with the competitive outcome. Monopolization of a competitive industry in these circumstances is a policy instrument available to the regulator which may be welfare improving.

## 2-3 REGULATED EQUILIBRIUM

The result illustrated in Figure 2-2-2 is an example of what is commonly termed "market failure". When this occurs, the government may decide to directly influence resource allocation by regulatory control of the industry. Before proceeding with detailed analysis of individual instruments, it must be noted that there are two necessary conditions for any regulatory control to be effective. In Chapter One Section 1-4, it was shown that these issues have begun to be considered in the literature with varying degrees of rigour.

Firstly, the control must be binding on the plans of the agent. With reference to Figure 2-2-2, any quota must be set at some output level less than the profit maximizing level $Q^C$. A nonbinding constraint has no effect whatsoever, ceteris paribus, on the behaviour of agents in the market.

Secondly, the binding control must also be enforced. Market equilibrium is derived on the basis of constrained maximization. At the competitive equilibrium point each individual consumer and producer has maximized utility and profits subject to their respective economic constraints. The introduction of a further binding constraint on any agent must necessarily reduce the value of his objective function. A rational maximizing agent will not therefore comply voluntarily with such an external control.

Compliance with a regulation requires inducement in the form of either a reward for so doing or punishment for violation. These alternatives are identical on efficiency criteria, differing only in their distributional effects [Coase, 1960]. Here it is assumed that the regulator punishes violations of the constraint by means of a monetary penalty, the form of which will be discussed with reference to each regulatory instrument separately.

The existence of a penalty of itself has no more effect on behaviour than the announcement of the control. The essential requirement is that individual agents perceive some likelihood of being punished if they violate the constraint. It is this probability-weighted expected penalty that influences agents' decisions. Given the existence of a regulation which embodies both of the above features, a regulated equilibrium can be defined.

A regulated equilibrium occurs when market supply and demand are equated within the regulatory environment and emerges as the result of competitive behaviour by individual agents. At a regulated equilibrium consumers are maximizing utility subject to their budget constraint and producers are maximizing profit subject to the technical and financial constraints on production. The only difference between this process and that described by (2-2-8) is that here individual maximization takes place within the additional behavioural constraints imposed by the regulatory regime. This clearly implies that in the absence of effective regulatory constraints, the unregulated competitive equilibrium will emerge.

The following analysis illustrates the derivation of a regulated equilibrium firstly in the case of a sales tax and secondly under an output quota.

## 2-4 THE EFFECT OF A SALES TAX ON THE INDIVIDUAL FIRM

Assume that a specific sales tax of \$t per unit is levied on the output of the externality-generating industry. The profit function of a representative firm in the industry now becomes

$$(2\text{-}4\text{-}1) \qquad \pi(q) = P.q - c(q) - t.q$$

Differentiating with respect to output gives the following first-order condition

$$(2\text{-}4\text{-}2) \qquad \frac{\partial \pi(q)}{\partial q} = P - c'(q) - t = 0$$

Comparing (2-2-4) and (2-4-1) it is evident that, with any strictly positive tax rate, the before-tax price received by the seller from the market must rise by the full amount of the tax in order for the same quantity to be supplied. That is, the supply curve of the individual firm shifts vertically by the amount of the tax. Alternatively at any given market price, with positive and strictly increasing marginal costs, the optimal output falls and the individual firm's supply curve shifts to the left.

The aggregation of the individual responses causes a reduction in market supply at any given price. Interaction with the demand curve results in an increase in market price and a decrease in market quantity, the precise measures of tax incidence and equilibrium output change being dependent on the relative price responsiveness of the market supply and demand curves. The first-best socially optimal output level, $\hat{Q}$ in Figure 2-2-2, can then be induced by imposing the tax rate $\hat{t}$

$$\hat{t} = MSB(\hat{Q}) - MPC(\hat{Q}) = \bar{P}_{\hat{Q}} - \underline{P}_{\hat{Q}} \tag{2-4-3}$$

which equates tax-inclusive marginal private cost of production with the demand price at this output level.

This is the traditional analysis of the incidence of a tax in a competitive market which shows that the first-best solution can be achieved through the use of an appropriate Pigouvian tax rate. The analysis however ignores many of the realities of taxation.

Under most taxation regimes, a taxpayer's initial tax liability is limited by the level of self declaration of taxable sales. Individuals base their declaration decisions on the effective tax rate which is defined here to be the expected tax rate. A deterrent to nonpayment is necessary to induce some degree of compliance with the regulation.

With no enforcement, the effective tax rate is zero and individual behaviour is unaffected. In this case the industry continues to operate at the competitive equilibrium irrespective of the announced tax rate.

Assume therefore that the announced tax rate of \$t per unit, payable on declared output x, is enforced by a monetary fine of \$F levied as a result of the detection of undeclared output with some probability $\rho$. Either or each of these components is some function of actual and declared output.

In the following analysis the industry is constrained to be in equilibrium so that all output produced, denoted by Q and q for the industry and representative firm respectively, is also sold and thus liable for tax. The tax-inclusive costs of production are now non-deterministic. There are of course a range of possible outcomes from undetected violations to full detection of undeclared output. Concentration here on the two extremes develops the essential aspects of the problem. These extreme outcomes are

(2-4-4) $\quad C_1(q,x) = c(q) + tx$

(2-4-5) $\quad C_2(q,x) = c(q) + tx + F(q,x)$

$C_1$ represents costs in the event of non-detection and occurs with probability $1-\rho$ while $C_2$ shows costs when any violation is punished which occurs with probability $\rho$. The expected cost of production at any level of output is therefore

(2-4-6) $\quad E[C(q,x)] = (1-\rho)C_1(q,x) + \rho C_2(q,x)$

$$= c(q) + tx + G(q,x)$$

substituting from (2-4-4) and (2-4-5) where

(2-4-7) $G(q,x) = \rho F(q,x)$ ; $G(0,x) = 0$ , $G(k,k) = 0$,

$G_q > 0$ , $G_x < 0$, where $G_i = \partial G(i,j)/\partial i$

$G_q$ is the marginal expected penalty with respect to actual output and $G_x$ is the marginal expected penalty with respect to declared output. It is assumed that the expected penalty increases with actual output at any given declaration level and decreases with declared output at any given production level.

Equation (2-4-7) gives the expected penalty as a product of its probability and monetary fine components. Individual producers are assumed to be risk neutral therefore only the absolute magnitude of the expected penalty is relevant to output decisions rather than its component parts. Thus a high probability/low fine combination generates an identical optimal output to a low probability/high fine combination which yields the same expected penalty.

No penalty is incurred for overreporting, truthful declarations, or at zero output. It must be noted however that profit maximization precludes overreporting and so in the analysis that follows

(2-4-8) $x \leq q$

The expected penalty as shown in (2-4-7) is a product of the monetary fine and the probability of detection and prosecution. The probability term is of necessity finite.

(2-4-9) $0 \leq \rho(q,x) \leq 1$

It is assumed here that the monetary fine is also finite. The justification for this assumption will be established in Section 2-7 where further discussion of the form of the expected penalty function occurs. If either of the component terms is zero, so is the product. Hence

(2-4-10) $$G(q,x)\Big|_{\rho(q,x)\,=\,0} = G(q,x)\Big|_{F(q,x)\,=\,0} = 0$$

The firm now acts to maximize profits within the regulatory framework. Expected profits for the competitive firm are given by

(2-4-11) $$\pi(q,x,\alpha) = p.q - c(q) - tx - G(q,x,\alpha)$$

where $\alpha$ is some parameter representing the structure of the expected penalty function which is exogenous to the firm.

(2-4-12) $$G_\alpha > 0$$

The firm chooses the output and declaration levels which maximize the value of (2-4-11). Expected profit is assumed to be locally concave jointly in declared and actual output. Differentiating with respect to the firm's decision variables gives the first-order conditions which are sufficient for a maximum.

(2-4-13) $$\frac{\partial\pi(q,x,\alpha)}{\partial q} = P - c'(q) - G_q \leq 0, \quad < \text{ if } q^* = 0$$

(2-4-14) $$\frac{\partial\pi(q,x,\alpha)}{\partial x} = -t - G_x \lesseqgtr 0 \quad , \begin{array}{l} < \text{ if } x^* = 0 \\ > \text{ if } x^* = q^* \end{array}$$

Solving these first-order conditions gives the firm's optimal choices of actual and declared output levels as functions of the market price, the unit rate of sales tax, and the parameter of the expected penalty function.

(2-4-15) $$q^* = q^*(P,t,\alpha) \quad , \; x^* = x^*(P,t,\alpha)$$

Equation (2-4-13) shows that, at the optimal output level, price equals marginal cost of production together with the effect of marginal changes in output on the expected penalty. If, however, the penalty-inclusive marginal cost always exceeds the market price, optimal output is zero.

Equation (2-4-14) shows that the optimal declaration level is found by comparing the unit tax rate with the marginal effect of declaration changes on the expected penalty. If a unit increase in declaration always reduces the expected penalty by a smaller amount than the unit tax rate, the optimal declaration is zero. When the reverse condition holds, full declaration is the optimal strategy.

Assuming that conditions sufficient to generate interior solutions to equations (2-4-13) and (2-4-14) hold, it is possible to examine the effect of parameter changes on the firm's behaviour by totally differentiating the first-order conditions. This gives

(2-4-16) $$dP - c''(q)dq - G_{qq}dq - G_{qx}dx - G_{q\alpha}d\alpha = 0$$

(2-4-17) $$-dt - G_{xq}dq - G_{xx}dx - G_{x\alpha}d\alpha = 0$$

Rewriting in matrix form gives

(2-4-18) $$\begin{bmatrix} G_{qq} + c''(q) & G_{qx} \\ G_{xq} & G_{xx} \end{bmatrix} \begin{bmatrix} dq \\ dx \end{bmatrix} = \begin{bmatrix} dP - G_{q\alpha}d\alpha \\ -dt - G_{x\alpha}d\alpha \end{bmatrix}$$

and rearranging

(2-4-19) $$\begin{bmatrix} dq \\ dx \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} G_{xx} & -G_{qx} \\ -G_{xq} & G_{qq} + c''(q) \end{bmatrix} \begin{bmatrix} dP - G_{q\alpha}d\alpha \\ -dt - G_{x\alpha}d\alpha \end{bmatrix}$$

where $\Delta$ is the determinant of the coefficient matrix and is strictly positive by the assumption of concavity of the profit function. Thus

(2-4-20) $\Delta = (G_{qq} + c''(q))\,G_{xx} - (G_{qx})^2 > 0$; using $G_{qx} = G_{xq}$

PROPOSITION 2-4-1: Suppose that an individual profit maximizing firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is concave:

(i) Supply increases with price.

(ii) Increases in price may increase or decrease the firm's optimal output declaration depending on the sign of the cross-derivative of the expected penalty function with respect to declared and actual output.

(iii) Increases in the tax rate may increase or decrease the optimal output of the firm at any given price for which supply is strictly positive, depending on the sign of the cross-derivative of the expected penalty function with respect to actual and declared output.

(iv) The effect of a change in price on declared output is equal and opposite to that of a change in the tax rate on actual output.

(v) Increases in the tax rate reduce the firm's optimal output declaration.

Proof:

(i) Using (2-4-19) with $dt = d\alpha = 0$

(2-4-21) $$\frac{dq^*}{dP} = \frac{G_{xx}}{\Delta} \gtreqless 0 \text{ if and only if } G_{xx} \gtreqless 0$$

Concavity of the firm's profit function requires overall convexity of the firm's penalty-inclusive cost structure as shown by (2-4-20). It also implies that the convexity in production cost dominates any possible non-convexity in output of the expected penalty function. Thus $G_{qq} + c''(q) > 0$ which from (2-4-20) ensures that $G_{xx}$ is strictly positive and the result holds.

(ii) Using (2-4-19) with $dt = d\alpha = 0$

(2-4-22) $$\frac{dx^*}{dP} = \frac{-G_{xq}}{\Delta} \gtreqless 0 \text{ if and only if } G_{xq} \lesseqgtr 0$$

where $G_{xq}$ is the cross-derivative of the expected penalty function and hence the result holds.

(iii) Using equation (2-4-19) with $dP = d\alpha = 0$

(2-4-23) $$\frac{dq^*}{dt} = \frac{G_{xq}}{\Delta} \gtreqless 0 \text{ if and only if } G_{xq} \gtreqless 0$$

and the result holds.

(iv) The result follows from (2-4-22) and (2-4-23).

(v) Using equation (2-4-19) with $dP = d\alpha = 0$

(2-4-24) $$\frac{dx^*}{dt} = \frac{-1}{\Delta}[G_{qq} + c''(q)] \lesseqgtr 0 \text{ if and only if } G_{qq} + c''(q) \gtreqless 0$$

From the assumption of concavity of the profit function $G_{qq} + c''(q) > 0$ and the result holds.

□

Results (i) and (v) of Proposition 2-4-1 are intuitively appealing. With a marginal expected penalty that increases in size with the severity of the violation, the convexity of the expected penalty function reinforces the convexity of production cost and hence profitable expansion in output requires an increasing market price. With respect to the other choice variable of the firm, a ceteris paribus increase in the tax rate raises the cost of any given output declaration. Faced with this situation profit maximization dictates that the firm's optimal declaration falls.

Result (iii) states that a change in price, which from part (i) of the Proposition serves to increase the optimal output of the firm at any given tax rate and penalty structure, may alter the firm's optimal declaration in either direction or not at all. The qualitative nature of the result depends on how the price-induced change in output affects the marginal expected penalty with respect to changes in declaration.

Equation (2-4-14) shows that with $0 < x^* < q^*$, the optimal declaration is found by balancing the marginal expected penalty $G_x$ with the tax rate. If the price-induced increase in output acts to increase the absolute size of this term at any given declaration, that is $G_{xq} < 0$, then, given $G_{xx} > 0$, the profit-maximizing strategy for the firm in accordance with equation (2-4-14) is to increase the size of its output declaration. The converse result is established by using the same reasoning with opposite assumptions concerning $G_{xq}$.

Result (ii) is also not immediately obvious. For instance, together with result (v), it suggests that an increase in the rate of sales tax can reduce the firm's optimal output declaration but raise actual output at any price leading to a rightward shift in its supply curve. This is directly opposite to the predictions of traditional tax analysis which arise from (2-4-2). To preclude this possibility, given the conditions sufficient to generate the other results in Proposition 2-4-1, it is necessary to assume that $G_{qx} = G_{xq} < 0$.

Finally, (2-4-22) and (2-4-23) show that the effect of a change in the tax rate on produced output is equal and opposite to that of a change in price on declared output. This type of cross-symmetrical result is commonly found and arises from maximization over a continuous surface which ensures that marginal responses to parameters are in effect equalized.

Figure 2-4-1 below gives a graphical presentation of these results for the representative firm. The diagram is for illustrative purposes only; the precise nature of the expected penalty function and its implications are discussed in Section 2-7.

As Figure 2-4-1 illustrates, actual output produced is a composite of declared and undeclared output. Panel (i) shows declared output against price while undeclared output is given in panel (ii). Panel (iii) compares declared output from panel (i), represented by the dashed line, with actual output in the regulated environment and pre-tax supply given by the solid line and dotted line respectively. All prices shown are tax-inclusive consumer prices and the curves are constructed using a given rate of sales tax $t_0 > 0$ and a given expected penalty function structure $\alpha_0$.

Figure 2-4-1: Firm supply with an enforced sales tax

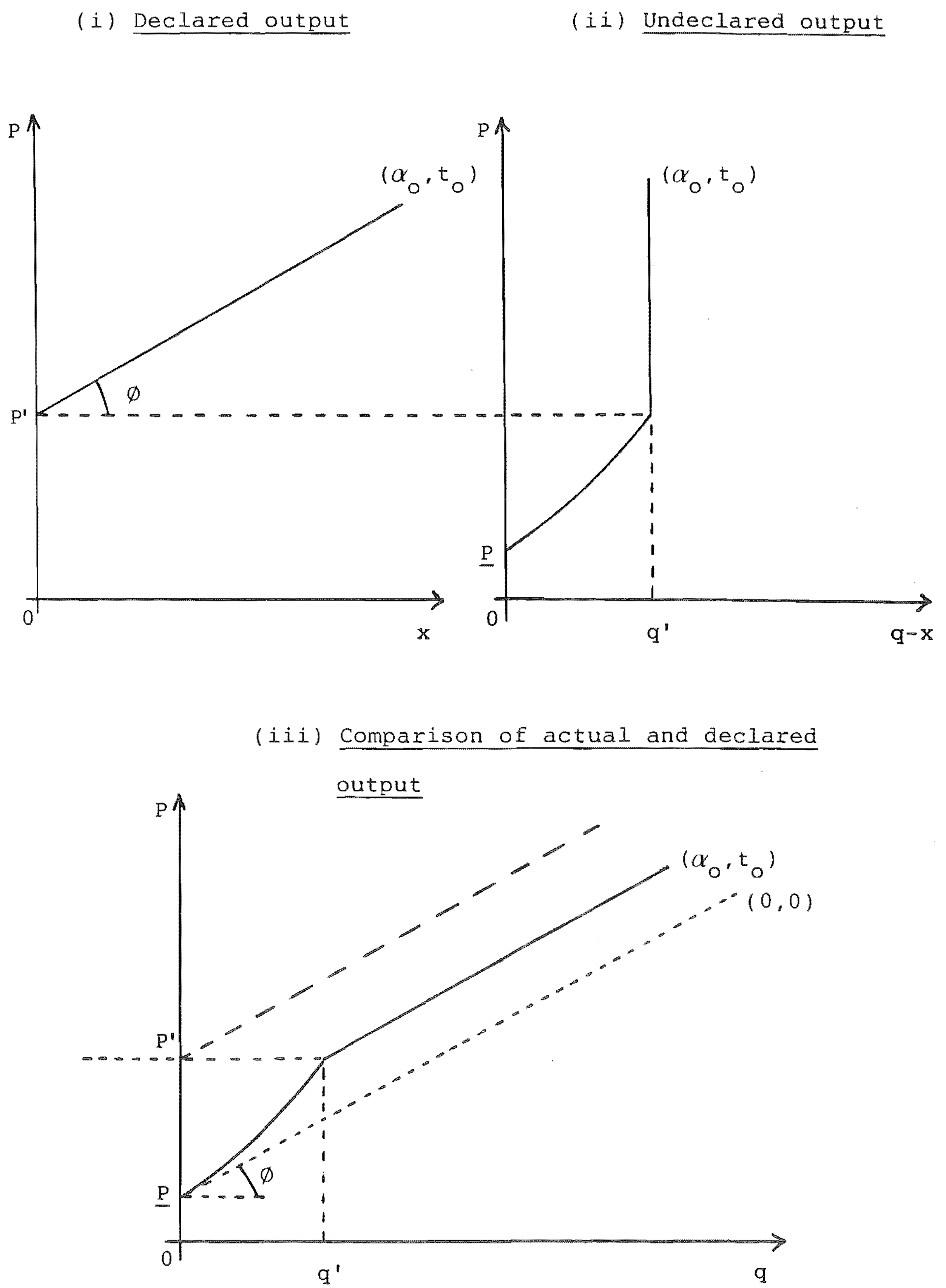


Figure 2-4-1 shows the initial q' units produced by the firm as all undeclared.

PROPOSITION 2-4-2: Suppose that an individual firm with strictly increasing marginal production cost operating in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty which does not totally prohibit undeclared output but is convex in output and the size of violation. At sufficiently low prices, the firm pays no taxes and sells all output illegally.

Proof.

Expressions (2-4-13) and (2-4-14) show the sizes of actual and declared output at any price level, given values of the tax rate and the expected penalty function, and hence also reveal the extent of undeclared output $q^* - x^*$.

The assumption that the expected penalty does not totally prevent undeclared output can be expressed as

$$(2\text{-}4\text{-}25) \qquad \lim_{q \to 0^+} G_x(q,0,\alpha^0) > -t^0$$

which from (2-4-14) results in zero declaration being the optimal strategy. Given (2-4-25) then, in the neighbourhood of $q^* = 0$, $q^* - x^* > 0$ and all output produced is undeclared.

Assuming that no output is declared, the penalty-inclusive marginal cost of producing an additional unit of undeclared output is

$$(2\text{-}4\text{-}26) \qquad MC(0) = c'(q) + G_q(q,0,\alpha^0)$$

Alternatively, the marginal cost of producing a unit of declared output is

$$(2\text{-}4\text{-}27) \qquad MC(q) = c'(q) + t^0$$

Comparing (2-4-26) and (2-4-27), profit maximization requires that if an additional unit is produced, it is undeclared for all units such that

$$(2\text{-}4\text{-}28) \qquad G_q(q,0,\alpha^0) < t^0$$

It is assumed that the expected penalty is convex in the size of undeclared output and, from (2-4-21), actual output increases with price. Assume then that there is some $P' > 0$ such that

$$(2\text{-}4\text{-}29) \qquad G_q(q^*(P',t^0,\alpha^0),0,\alpha^0) = t^0$$

Using (2-4-25) and (2-4-28) therefore, for all prices $P < P'$, $x^* = 0$, no taxes are paid and any output that the firm produces is sold illegally.

□

The result of Proposition 2-4-2 is appealing. If there was no enforcement of the tax, the supply price of declared output would exceed that of undeclared output at any quantity by the tax rate $t_0$. Assuming consumers face no penalty,[5] arbitrage dictates that there be one market price irrespective of whether output is legal or illegal and hence no output will be declared [Smith, 1976]. Neither will output be declared if the marginal expected penalty with respect to output, given that no output is declared, is always less than the tax rate. This implies that high enough taxes with a given expected penalty structure will make criminals of all producers.

Figure 2-4-1 portrays the particular case where, for prices exceeding $P'$, all additional output produced is declared and hence undeclared output remains constant at $q'$.

In general this will not occur.

PROPOSITION 2-4-3: Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is strictly concave. An increase in market price, from a level at which both actual and declared output are strictly positive, increases, decreases, or leaves unchanged the extent of undeclared output, as the effect on the marginal expected penalty with respect to declared output of an increase in output declaration exceeds, is less than, or equals the negative of that of an increase in actual output.

Proof.

$$\text{(2-4-30)} \qquad \frac{d}{dP}(q^* - x^*) = \frac{dq^*}{dP} - \frac{dx^*}{dP}$$

$$= \frac{G_{xx}}{\Delta} - \left(\frac{-G_{xq}}{\Delta}\right)$$

using (2-4-21) and (2-4-22). Therefore

$$\text{(2-4-31)} \qquad \frac{d}{dP}(q^* - x^*) \gtreqless 0 \text{ if and only if } G_{xx} \gtreqless -G_{xq}$$

□

Another result which is closely related to this concerns the possibility of different degrees of price responsiveness in actual output in the regulated and unregulated frameworks.

PROPOSITION 2-4-4: Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is concave. In the region where both declared output and actual output are strictly positive, the firm's supply curve in the presence of the sales tax is less price responsive than the untaxed supply curve if the expected penalty function is convex.

Proof.

Comparing (2-2-6) and (2-4-21) shows that $\frac{dq^*}{dP} \gtrless \frac{dq^s}{dP}$ if and only if

$$\frac{G_{xx}}{(G_{qq} + c''(q))G_{xx} - (G_{xq})^2} \gtrless \frac{1}{c''(q)} \tag{2-4-32}$$

where q* represents supply in the presence of the sales tax and $q^s$ represents untaxed supply. Simplifying and rearranging (2-4-32) yields the result that $\frac{dq^*}{dP} \gtrless \frac{dq^s}{dP}$ if and only if

$$G_{qq}G_{xx} - (G_{xq})^2 \lessgtr 0 \tag{2-4-33}$$

Expression (2-4-33) is a determinant of the matrix of second partial derivatives of the expected penalty function G. If G is convex, (2-4-33) has a positive sign and the taxed supply curve is less price responsive than untaxed supply.

□

This result is intuitively appealing. The convexity of the expected penalty function reinforces the convexity already present in production cost. Any given increase in output therefore requires a larger increase in price than in the unregulated environment and hence supply is less price responsive with the sales tax than without.

The effect of a change in the tax rate on a firm's behaviour can also be examined. In standard tax analysis the effect would be to alter the supply price at any given quantity by the change in the tax rate leading to a vertical shift in the supply curve. Here, where tax violations are allowed for, the precise nature of the effect of a change in the tax rate on the firm's supply curve depends on the properties of the expected penalty function.

PROPOSITION 2-4-5: Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is concave. Assuming that the cross-derivative of the expected penalty function with respect to actual and declared output is negative and that the expected penalty is convex in actual output, an increase in the tax rate increases the amount of output that is produced illegally before any is declared.

Proof.

From (2-4-29) the maximum level of output that is produced before any is declared occurs where the marginal expected penalty with respect to output, given that no output is declared, is equal to the tax rate. In Figure 2-4-1 this output level is given by q' in panel (ii).

If the tax rate is increased to some $t' > t^0$ then

(2-4-34) $$G_q(q',0,\alpha^0) < t'$$

Given the assumption that $G_{qq} > 0$, equality between the marginal expected penalty with respect to output, given that no output is declared, and the tax rate, is restored at some $q'' > q'$. Therefore

(2-4-35) $$\frac{d}{dt}\left[ q_{max}\Big|_{x^*=0} \right] > 0$$

and the result holds.

□

Alternatively the result can be demonstrated by examining the profit-maximizing declaration strategy at any output level and tax rate. At the original tax rate $t^0$ and output level q' in Figure 2-4-1

(2-4-36) $$\lim_{x \to 0^+} G_x(q',x,\alpha^0) = -t^0$$

which results from the equality condition in (2-4-14).

If the tax rate is increased to some $t' > t^0$ then

(2-4-37) $$G_x(q',0,\alpha^0) > -t'$$

Given the assumption that $G_{xq} < 0$, equality between the marginal expected penalty with respect to declared output, given that no output is declared, and the tax rate, is restored at some $q'' > q'$. This approach then also generates the result presented in (2-4-35) and hence proves the Proposition.

Equations (2-4-23) and (2-4-24) ensure that, provided the firm's profit function is strictly concave and $G_{xq} < 0$, an increase in the tax rate will lower both declared output and actual output at any given price. From (2-4-35) it may be expected that an increase in the tax rate will also lead to an increase in undeclared profit. This is not in general so, however.

PROPOSITION 2-4-6: Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is concave. At any price level for which both actual and declared outputs are strictly positive, an increase in the tax rate increases, decreases, or leaves unchanged, the extent of undeclared output, if and only if the combined effect on the marginal expected penalties with respect to declared output and actual output of a change in actual output exceeds, is less than, or equals, the negative of the effect of a change in actual output on marginal production cost.

Proof.

$$\frac{d}{dt}(q^* - x^*) = \frac{dq^*}{dt} - \frac{dx^*}{dt} \tag{2-4-38}$$

Substituting from (2-4-23) and (2-4-24)

$\frac{d}{dt}(q^* - x^*) \gtreqless 0$ if and only if

$$\frac{G_{qx}}{\Delta} - \frac{-[G_{qq} + c''(q)]}{\Delta} \gtreqless 0 \tag{2-4-39}$$

Rearranging and using $G_{qx} = G_{xq}$

$\frac{d}{dt}(q^* - x^*) \gtrless 0$ if and only if

$$G_{xq} + G_{qq} \gtrless -c''(q) \qquad (2\text{-}4\text{-}40)$$

where the left-hand side of (2-4-40) is the combined effect of a change in actual output on the marginal expected penalties with respect to declared output and actual output respectively, and the right-hand side is the effect of a change in actual output on marginal production cost.

□

If ">" holds in (2-4-40), an increase in the tax rate will increase the extent of undeclared output. This is an intuitively appealing result but, as shown in Proposition 2-4-6, is by no means a necessary consequence of enforcement. It is possible that the magnitude of the effect of a change in output on the marginal expected penalty with respect to declared output exceeds that of its effect on the marginal expected penalty with respect to actual output. This could be so if the structure of the expected penalty function is designed to encourage more truthful declaration as output levels rise. In this case, $G_{xq} + G_{qq} < 0$. The direction of the inequality in (2-4-40) then depends on the degree of convexity of the firm's penalty-exclusive production cost.

Propositions 2-4-1 to 2-4-6 are based on the assumption of a given expected penalty function exogenous to the firm. Changes in this function, however, through the parameter $\alpha$, will affect the firm's behaviour.

<u>PROPOSITION 2-4-7</u>: <u>Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable</u>

on declared output and enforced by means of an expected monetary penalty, that the cross-derivative of the expected penalty function with respect to actual and declared outputs is negative, and that the firm's profit function is concave. There is an equivalence between the effects on actual and declared outputs of a change in the expected penalty and those of changes in taxes and price. An increase in the expected penalty reduces actual output if the effect on the marginal expected penalty with respect to actual output of the change in penalty is sufficiently large whereas it increases declared output if the effect of the change in penalty on the marginal expected penalty with respect to actual output is sufficiently small.

Proof.

Using (2-4-19) with dp = dt = 0

$$(2\text{-}4\text{-}41) \qquad \frac{dq^*}{d\alpha} = -\frac{G_{xx}}{\Delta} G_{q\alpha} + \frac{G_{qx}}{\Delta} G_{x\alpha}$$

Substituting from (2-4-21) and (2-4-23) gives

$$(2\text{-}4\text{-}42) \qquad \frac{dq^*}{d\alpha} = \frac{dq^*}{dt} G_{x\alpha} - \frac{dq^*}{dP} G_{q\alpha}$$

which shows the equivalence between the effects, on actual output, of a change in the expected penalty parameter and those of changes in taxes and price.

From (2-4-42) $\frac{dq^*}{d\alpha} \gtrless 0$ if and only if

$$(2\text{-}4\text{-}43) \qquad G_{q\alpha} \lesseqgtr [\frac{dq^*}{dt}/\frac{dq^*}{dP}]G_{x\alpha}$$

Given the assumptions that $G_{xq} < 0$ and $G_{x\alpha} < 0$, the right-hand side of (2-4-43) is some positive number. From (2-4-43) then, increases in penalties reduce actual output if and only if $G_{q\alpha}$ is sufficiently large.

Again from (2-4-19), with dp = dt = 0

$$(2\text{-}4\text{-}44) \qquad \frac{dx^*}{d\alpha} = \frac{G_{xq}}{\Delta} G_{q\alpha} - \frac{[(G_{qq} + c''(q)]}{\Delta} G_{x\alpha}$$

Substituting from (2-4-22) and (2-4-24) gives

$$(2\text{-}4\text{-}45) \qquad \frac{dx^*}{d\alpha} = \frac{dx^*}{dt} G_{x\alpha} - \frac{dx^*}{dP} G_{q\alpha}$$

which shows the equivalence between the effects, on declared output, of a change in the expected penalty parameter and those of changes in taxes and price.

Rearranging (2-4-45) reveals that $\frac{dx^*}{d\alpha} \gtreqless 0$ if and only if

$$(2\text{-}4\text{-}46) \qquad G_{q\alpha} \lesseqgtr [\frac{dx^*}{dt}/\frac{dx^*}{dP}]G_{x\alpha}$$

Given the assumptions that $G_{xq}$ and $G_{x\alpha}$ are negative, the right-hand side of (2-4-46) is some positive number. From (2-4-46) then, increases in penalties increase declared output if and only if $G_{q\alpha}$ is sufficiently small.

□

Intuitively, an increase in the structure of the penalty function may be expected to raise the marginal expected penalty cost of output expansion at any given declaration level and raise the marginal expected penalty saving of increased declaration at any given actual output level where undeclared output is produced. This is the reasoning behind the assumption that $G_{q\alpha} > 0$ and $G_{x\alpha} < 0$.

A result which is widely held in the deterrence and crime literatures, both theoretical and empirical, is that increases in expected penalties reduce the extent of illegal behaviour.[6] This is now investigated in the present context of tax declarations.

PROPOSITION 2-4-8: Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty and that the firm's profit function is concave. Assuming that the cross-derivative of the expected penalty function with respect to actual and declared outputs is negative, that the effect of an increase in the expected penalty is to increase the absolute values of the marginal expected penalties with respect to actual and declared output, and that the expected penalty function is convex in output, an increase in the expected penalty reduces the amount of output that is produced illegally before any is declared.

Proof.

In Proposition 2-4-2 the maximum amount of output produced before any is declared is given by the value of q that satisfies (2-4-29) given $t^0$ and $\alpha^0$. This is taken to occur, as portrayed in Figure 2-4-1, at output level q' and price P'. Assuming that $G_{q\alpha} > 0$ then at some $\alpha' > \alpha^0$

$$(2\text{-}4\text{-}47) \qquad G_q(q^*(P',t^0,\alpha'),0,\alpha') > t^0$$

Given that $G_{qq} > 0$, the equality between the marginal expected penalty with respect to output, given that no output is declared, and the tax rate, is restored at some $q'' < q'$ which from (2-4-21) occurs at some $P'' < P'$. Therefore

$$(2\text{-}4\text{-}48) \qquad \frac{d}{d\alpha}\left[ q_{max} \Big|_{x^*=0} \right] < 0$$

and the result holds.

□

Alternatively, as for Proposition 2-4-5, the result can be demonstrated by examining the profit-maximizing declaration strategy at any output level and tax rate. Using (2-4-36) and assuming that $G_{x\alpha} < 0$

$$(2\text{-}4\text{-}49) \qquad \lim_{x \to 0^+} \quad G_x(q',x,\alpha') < -t^0$$

Given that $G_{xx} > 0$ (2-4-14) implies that the optimal declaration is strictly positive at the output level $q'$ and penalty parameter $\alpha = \alpha'$.

With $G_{xq} < 0$, equality between the marginal expected penalty with respect to declared output, given that no output is declared, and the tax rate, is restored at some $q'' < q'$. This approach then also generates the result presented in (2-4-48) and hence proves the Proposition.

This result is consistent with the thesis that increased penalties reduce illegal behaviour. However, it does not of itself imply that the extent of undeclared output, at any price level for which both declared and actual output are strictly positive, will be reduced by an increase in the expected penalty.

It may be expected that an increase in the expected penalty at any given level of illegal activity would encourage a more truthful declaration of production activities and thus $dx^*/d\alpha > 0$. Assuming that this is the case, the extent of undeclared output unambiguously decreases if the increase in penalty reduces the level of actual output. This is sufficient to ensure the result but is not necessary. A necessary and sufficient condition to ensure greater compliance with the tax is that the effect on declared output of the increase in expected penalty exceeds that on actual output.

<u>PROPOSITION 2-4-9</u>: <u>Suppose that an individual firm in a competitive market is faced with a strictly positive unit sales tax payable on declared output and enforced by means of an expected monetary penalty, that the cross-derivative of the expected penalty with respect to declared and actual output is negative, and that an increase in the expected penalty acts to increase the absolute values of the marginal expected penalties with respect to actual and declared output. There is an equivalence between the effect on undeclared output of a change in the expected penalty and that of a change in taxes and price. Higher penalties reduce the extent of undeclared output if it increases with the tax rate and this</u>

<u>increase outweighs any possible decrease that occurs with an increase in price</u>.

<u>Proof</u>.

Using (2-4-42) and (2-4-45)

$$\frac{d}{d\alpha}(q^*-x^*) = \frac{dq^*}{d\alpha} - \frac{dx^*}{d\alpha} \gtrless 0 \text{ if and only if} \tag{2-4-50}$$

$$\frac{d}{dt}(q^*-x^*)G_{x\alpha} - \frac{d}{dP}(q^*-x^*)G_{q\alpha} \gtrless 0 \tag{2-4-51}$$

From (2-4-50) and (2-4-51), a reduction in the extent of undeclared output as a result of an increase in the expected penalty requires that "<" holds in (2-4-51).

Given that the extent of undeclared output increases with the tax rate, and that $G_{x\alpha} < 0$ and $G_{q\alpha} > 0$, (2-4-51) is unambiguously negative if the extent of undeclared output increases with price. If however undeclared output decreases in price then from (2-4-51)

$$\frac{d}{dt}(q^*-x^*) > \frac{d}{dP}(q^*-x^*)\frac{G_{q\alpha}}{G_{x\alpha}} \tag{2-4-52}$$

in order that undeclared output decreases as expected penalties are increased.

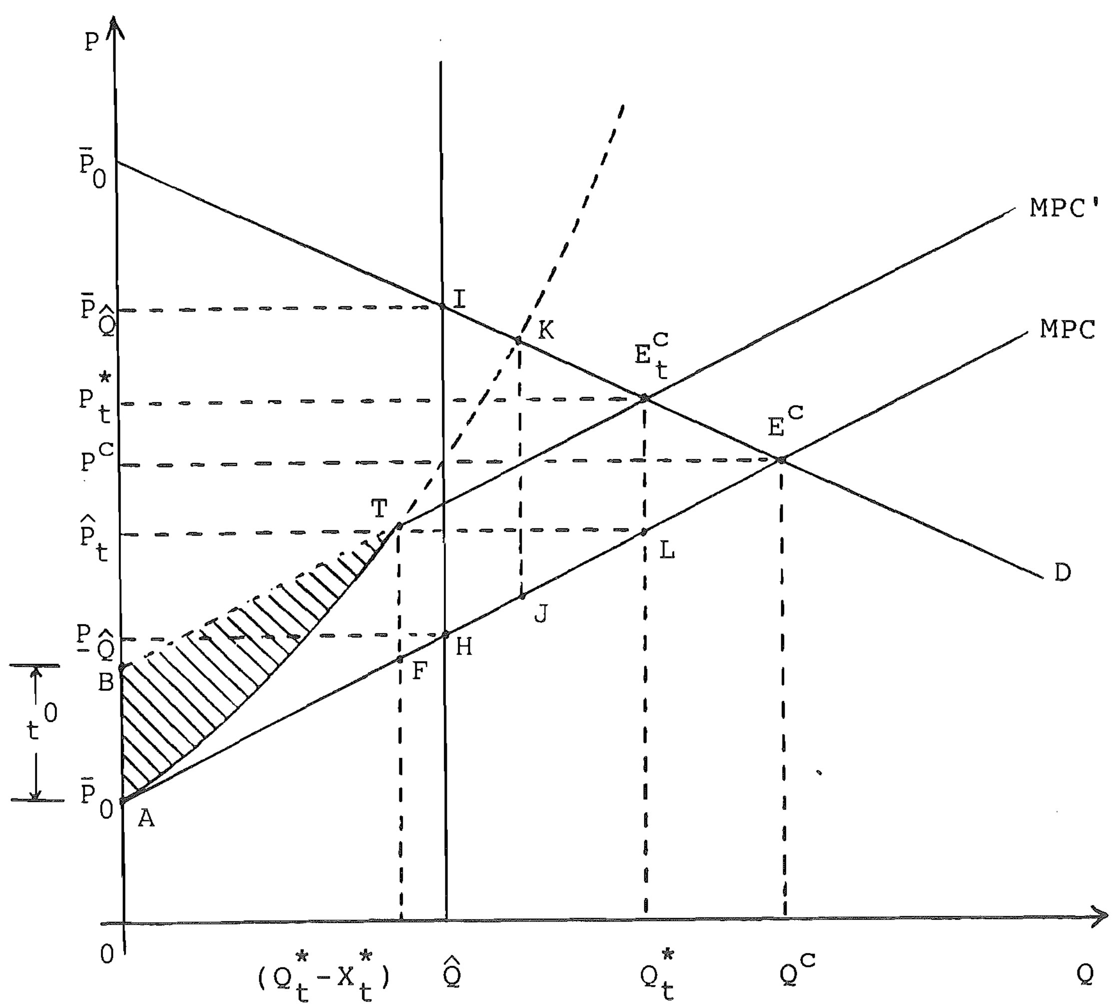
□

Propositions 2-4-1 to 2-4-9 describe aspects of how the behaviour of an individual firm is affected by the introduction of an enforced sales tax given that the firm's profit function is concave. Many of the intuitively appealing results which are widely held in similar form in the deterrence and crime literatures are present here but depend on the properties of the expected penalty function. These results are further discussed in Section 2-7 where different forms of the expected penalty function are examined.

## 2-5 THE EFFECT OF A SALES TAX ON THE INDUSTRY

As the behaviour of individual firms at any given price is altered by the existence of a sales tax and expected penalty function for false output declaration, so the competitive equilibrium of the industry will be affected leading to the regulated competitive equilibrium. This is shown in Figure 2-5-1 below.

Figure 2-5-1: Industry equilibrium with an enforced sales tax




Curves MPC and MPC' are constructed on the assumption that there is technical and economic efficiency within the industry so that, with identical firms, they differ only in aggregation scale from those of the individual firm in Figure 2-4-1. All relevant curves are drawn for a given fixed value of the expected penalty function parameter $\alpha = \alpha_0$.

The initial competitive equilibrium at $E^C$ occurs at an output $Q^C$ which exceeds the socially optimal level $\hat{Q}$ from Figure 2-2-2. A unit tax of $t^0$ is imposed which under traditional tax analysis would lead to a vertical shift in the industry supply curve of AB = FT.

Allowing for evasion of the tax, an illustrative expected penalty function gives a penalty-inclusive marginal cost curve ATK. From (2-4-14) the aggregate optimal amount of undeclared output is given by $Q_t^* - X_t^*$ at point T where the marginal expected penalty, given that no output is declared, equals the tax rate.

The regulated competitive industry equilibrium occurs at some point such as $E_t^C$ where the marginal expected penalty is equal to the difference between the demand price and private marginal production cost evaluated at that quantity which may or may not equal the tax rate. Thus at the regulated competitive equilibrium

(2-5-1) $$G_q(q_t^*, x^*, \alpha^0) = G_Q(Q_t^*, X^*, \alpha^0) = P_t^* - \hat{P}_t \leq t^0$$

where $P_t^* - \hat{P}_t$ is the difference between the demand price and private marginal production cost at the regulated equilibrium quantity $Q_t^*$, and

(2-5-2) $$0 \leq Q_t^* \leq Q^C \;;\; Q_t^* = zq_t^*$$

where z represents the number of firms in the industry.

Expressions (2-5-1) and (2-5-2) show that there are several possibilities at the regulated competitive equilibrium.

Firstly, from equation (2-5-1), the tax rate may or may not be binding on behaviour. Figure 2-5-1 illustrates a situation where it is, but, if the tax rate exceeded amount JK, behaviour would be solely determined by the expected penalty function and the rate of tax would be irrelevant. A binding tax rate requires an expected penalty function such that the marginal expected penalty with respect to actual output, given that no output is declared, exceeds the tax rate at some output level which does not exceed the regulated competitive equilibrium level. Thus, assuming a marginal expected penalty that is non-decreasing in output, the tax rate is binding on behaviour if

(2-5-3) $\quad G_Q(Q'X,\alpha^0) \geq t$ for some $Q' \leq Q^*_t$

Secondly, from expression (2-5-2) the regulated competitive equilibrium is bounded from above and below. It may coincide with the unregulated competitive equilibrium. For this to occur with a non-negative expected penalty function either the tax rate must be zero or any strictly positive tax rate must be accompanied by a zero expected penalty. Alternatively it is possible to regulate the industry out of existence.

In Figure 2-5-1, $\overline{P}_0 - \underline{P}_0$ represents the difference between demand price and marginal private cost of production for the initial unit of output from the industry. The regulated equilibrium occurs at zero output if either of the following circumstances eventuates.

Given $\overline{P}_0 - \underline{P}_0$, no output will be produced in the regulated environment if $t = \overline{P}_0 - \underline{P}_0$ and the expected penalty is such that the marginal unit of output is declared. This requires

(2-5-4) $$\operatorname{Lim}_{Q\to 0^+} G_Q(Q,X,\alpha^0) \geq t$$

If however the tax rate is $t > \overline{P}_0 - \underline{P}_0$ then the expected penalty function must be such that

(2-5-5) $$\operatorname{Lim}_{Q\to 0^+} G_Q(Q,X,\alpha^0) \geq \overline{P}_0 - \underline{P}_0 .$$

In this case it is the marginal expected penalty at zero output which is binding and causes the regulated competitive equilibrium to emerge at this level. A regulated competitive equilibrium at zero output cannot be generated with any tax rate $t < \overline{P}_0 - \underline{P}_0$ irrespective of the penalty used.

Finally, the regulated competitive equilibrium may occur at some intermediate point such as that shown on Figure 2-5-1. This requires either a binding tax rate $t < \overline{P}_0 - \underline{P}_0$ or any non binding tax rate with an expected penalty function such that

(2-5-6) $$\operatorname{Lim}_{Q\to 0^+} G_Q(Q,X,\alpha^0) < \overline{P}_0 - \underline{P}_0$$

This is the interesting case and is that most extensively treated in the following analysis.

From Figure 2-5-1 it appears that the regulated competitive equilibrium for a strictly positive tax rate and expected penalty occurs at an increased market equilibrium price and reduced market quantity compared with the unregulated competitive equilibrium.

PROPOSITION 2-5-1: Suppose that an industry which comprises a fixed and finite number of identical competitive firms subject to a unit sales tax, payable on declared output and enforced by means of an expected monetary penalty such that the firms' profit functions are concave, faces a market demand curve with non-zero finite price elasticity. Assuming that an increase in output increases the absolute value of the marginal expected penalty with respect to declared output, that the expected penalty function is convex in actual output, and that an increase in the expected penalty sufficiently raises the marginal expected penalty with respect to actual output;

(i) An increase in the tax rate raises the market equilibrium price and reduces industry equilibrium output only if the tax rate is binding at the initial regulated industry equilibrium.

(ii) An increase in the expected penalty at any given declaration and output in general raises the market equilibrium price and reduces industry equilibrium output. Suppose however that the effect on the marginal expected penalty with respect to declared output of an increase in declared output is equal and opposite to that of an increase in actual output and equal to the effect of a change in actual output on the marginal expected penalty with respect to actual

<u>output. In this case, if the marginal expected penalty with respect to output, given that no output is declared, is not binding at the initial regulated equilibrium, a change in the expected penalty has no effect on market equilibrium quantity or price</u>.

<u>Proof</u>.

From (2-4-15), (2-2-2) and (2-2-7) it is evident that market equilibrium price is a function of the tax rate and the expected penalty. Thus in industry equilibrium

(2-5-7) $\quad D(P(t,\alpha)) = S(P(t,\alpha),t,\alpha)$

where S(.) is the regulated industry supply function defined as the summation of individual firm supply curves in the regulated environment as for (2-2-7) in the unregulated environment.

There are two possibilities for the regulated industry equilibrium $Q_t^*$. Firstly,

(2-5-8) $\quad G_Q(Q_t^*,0,\alpha) = \bar{P}_{Q_t^*} - \underline{P}_{Q_t^*} < t.$

Here the regulated equilibrium occurs at some quantity $Q_t^*$ where the marginal expected penalty with respect to output, given that no output is declared, is equal to the difference between the demand price and private marginal production cost at that output level. The tax rate exceeds the value and hence does not directly affect behaviour. In Figure 2-5-1, this would occur if the tax rate exceeded amount JK.

Alternatively,

(2-5-9) $\quad G_Q(Q_t^*,0,\alpha) > \bar{P}_{Q_t^*} - \underline{P}_{Q_t^*} = t.$

In this case the marginal expected penalty with respect to output, given that no output is declared, exceeds the tax rate at the regulated industry equilibrium $Q_t^*$. This is the situation illustrated in Figure 2-5-1 at $E_t^c$. Here, increases in the expected penalty function may not affect the regulated industry equilibrium.

(i) Totally differentiating (2-5-7) gives

$$(2\text{-}5\text{-}10)\quad \frac{\partial S(P(t,\alpha),t,\alpha)}{\partial P}\frac{\partial P(t,\alpha)}{\partial t} + \frac{\partial S(.)}{\partial t} + \frac{\partial S(.)}{\partial P}\frac{\partial P(t,\alpha)}{\partial \alpha} + \frac{\partial S(.)}{\partial \alpha} - D'(P(t,\alpha))\frac{\partial P(t,\alpha)}{\partial t} - D'(P(t,\alpha))\frac{\partial P(t,\alpha)}{\partial \alpha} = 0$$

Rearranging with $\partial\alpha = 0$ gives

$$(2\text{-}5\text{-}11)\quad \frac{\partial P^*(t,\alpha)}{\partial t} = \frac{\partial S(.)/\partial t}{D'(P(t,\alpha))-\partial S(.)/\partial P} \gtrless 0 \text{ if and only if } \partial S(.)/\partial t \lessgtr 0$$

If (2-5-8) holds, $\partial S(.)/\partial t = 0$ and from (2-5-11) market equilibrium price is unaffected by changes in the tax rate. If this is not the case however, then, given $G_{xq} < 0$, an increase in the tax rate reduces individual firm supply at any price level by (2-4-33), and hence $\partial S(.)/\partial t < 0$. From (2-5-11) therefore, an increase in a binding tax rate raises market equilibrium price. Noting from (2-5-7) that

$$(2\text{-}5\text{-}12)\quad Q_t^* = D(P^*(t,\alpha))$$

where $Q_t^*$ is the regulated industry equilibrium output level

$$(2\text{-}5\text{-}13)\quad \frac{\partial Q_t^*}{\partial t} = D'(P(t,\alpha))\frac{\partial P^*(t,\alpha)}{\partial t} \gtrless 0 \text{ if and only if } \frac{\partial P^*(t,\alpha)}{\partial t} \lessgtr 0$$

Using (2-5-11) then, an increase in a binding tax rate reduces market equilibrium output in the regulated industry.

(ii) Rearranging (2-5-10) with $\partial t = 0$ gives

$$(2\text{-}5\text{-}14) \quad \frac{\partial P^*(t,\alpha)}{\partial \alpha} = \frac{\partial S(.)/\partial \alpha}{D'(P(t,\alpha)) - \partial S(.)/\partial P} \gtrless 0 \text{ if and only if } \frac{\partial S(.)}{\partial \alpha} \lessgtr 0$$

If (2-5-9) does not hold, then, given $G_{q\alpha}$ is sufficiently positive, an increase in the expected penalty reduces individual firm supply by (2-4-42), and hence $\partial S(.)/\partial \alpha < 0$. From (2-5-14) therefore, an increase in a binding expected penalty raises market equilibrium price. Using (2-5-12)

$$(2\text{-}5\text{-}15) \quad \frac{\partial Q_t^*}{\partial \alpha} = D'(P(t,\alpha)) \frac{\partial P^*(t,\alpha)}{\partial \alpha} \gtrless 0 \text{ if and only if } \frac{\partial P^*(t,\alpha)}{\partial \alpha} \lessgtr 0$$

Substituting from (2-5-14) then, an increase in a binding expected penalty reduces market equilibrium output in the regulated industry.

This result is true in general, given the assumptions of concavity of the individual firm's profit function and convexity in output of the expected penalty function, even if the marginal expected penalty, given that no output is declared, exceeds the tax rate at the regulated industry equilibrium.

Using (2-4-19), for any given expected penalty function and tax rate, there is a maximum amount of output initially produced before any is declared. At this output level the marginal expected penalty with respect to output, given that no output is declared, is equal to the tax rate. This gives a certain amount of undeclared output. From (2-4-33), however, the optimal extent of undeclared output is, in general, not independent of the output level of the firm, and hence the industry.

An increase in the expected penalty function will, from Proposition 2-4-8, reduce the amount of output that is initially produced before any is declared. If the optimal extent of undeclared output is not independent of the level of actual output, then, given that $G_{q\alpha}$ is sufficiently positive and using (2-5-14) and (2-5-15), the increase in expected penalty raises the market equilibrium price and lowers industry equilibrium output even though (2-5-9) holds.

If however (2-5-9) holds and $G_{qq} = G_{xx} = -G_{xq}$ such that $G_{qq} > 0$ and hence the individual firm's profit function is concave, from (2-4-33) the optimal extent of undeclared output is independent of actual output and determined by the output level where the marginal expected penalty with respect to output, given that no output is declared, is equal to the tax rate. In this case, therefore, individual firm and hence industry supply at the regulated equilibrium price level is unaffected by an increase in the expected penalty and so from (2-5-14) and (2-5-15) the increase in expected penalty leaves the regulated industry equilibrium unchanged.

$\square$

Equations (2-5-11) and (2-5-14) show that an increase in the expected penalty function or tax rate, which acts to decrease individual firm and hence industry supply at every price, increases the market equilibrium price. Multiplying this result by the price responsiveness of the demand curve in (2-5-12) and (2-5-15) gives the effect of the increase in the expected penalty function and tax rate respectively on the equilibrium quantity in the market. Given the new market equilibrium price, the individual firm's optimal regulated output can be found by solving (2-4-13) at this price which gives $q_t^*$ consistent with (2-5-2).

Following Proposition 2-5-1 and the discussion of the conditions required to generate each of the possible outcomes in (2-5-2) it is evident that, with an appropriate mix of the tax rate and the expected penalty, it is possible to restrain industry output to the socially optimal level $\hat{Q}$ in Figure 2-5-1. This can be accomplished by setting the tax rate and expected penalty so that either

(2-5-16) $\quad t = \hat{t} \; ; \; G_Q(\hat{Q},0,\alpha^0) \geq \bar{P}_{\hat{Q}} - \underline{P}_{\hat{Q}}$ or

(2-5-17) $\quad t > \hat{t} \; ; \; G_Q(\hat{Q},0,\alpha^0) = \bar{P}_{\hat{Q}} - \underline{P}_{\hat{Q}}$

Expression (2-5-16) describes the situation where the tax rate is set equal to the marginal externality at $\hat{Q}$ and the expected penalty is sufficient to ensure the declaration of at least the marginal unit of output. Alternatively (2-5-17) shows the case where the tax rate is prohibitive at this output level but the marginal expected penalty is such that $\hat{Q}$ is the regulated competitive equilibrium. Whether or not this occurs and which method is employed depends on the objectives of the regulator. These are discussed in Chapter Three.

While Proposition 2-5-1 shows that changes in a non-binding expected penalty or tax rate will not alter market equilibrium price and quantity, these changes will affect the size of the aggregate gain in industry profitability from tax evasion shown by area ABT in Figure 2-5-1. This point, which has implications for the policy mix used in regulating the industry, will also be discussed in Chapter Three.

## 2-6 THE EFFECTS OF AN OUTPUT QUOTA

Assume that instead of being regulated by a 'price-oriented' control such as the sales tax in Sections 2-4 and 2-5, firms in the externality-producing industry are now subject to quantitative restrictions. Denoting the unregulated equilibrium by $Q^c$ from Figure 2-2-2, the industry is subject to a strictly binding quota of R units.

(2-6-1) $R < Q^c$

This aggregate quota is tradeable in a competitive manner among firms within the industry. The assumption of identical profit maximizing firms ensures that, regardless of the initial allocation mechanism, in equilibrium the representative firm in the industry is subject to a quota of r units

(2-6-2) $r = R/z < q^c$

where z is the number of firms in the industry and $q^c$ is the individual firm's unregulated equilibrium output level.

At the equilibrium trading price of the quota, any firm holding more than r units of quota would lower the expected penalty cost on its marginal unit of output but would be reducing its profit through the opportunity cost of holding the excessive quota. Alternatively, any firm holding less than r units of quota may have a lower opportunity cost of holding the quota but, with an expected penalty that increases in the size of the constraint violation, profits will be reduced by the increased expected penalty cost that it incurs at any given output level with the smaller quota holding. Expression (2-6-2) therefore gives the profit maximizing level of quota holding for the individual firm.

The initial allocation of quota then determines the distribution of income within the industry but total industry income is dependent only on its aggregate size.[7]

As in the sales tax case, the imposition of an additional constraint on behaviour necessarily reduces profits. In the absence of deterrents to noncompliance, individual profit-maximizing behaviour will result in the competitive equilibrium outcome and therefore enforcement is required to make the control effective. Enforcement is carried out, here as before, by means of an expected monetary penalty $H(q,r,\beta)$ which exhibits similar properties to the expected penalty function $G(q,x,\alpha)$ in the sales tax case, thus

(2-6-3) $$H(q,r,\beta) = \rho F(q,r,\beta) \; ; \; H(0,r,\beta) = 0; \; H(k,k,\beta) = 0, \; k \geq 0; \quad H_q > 0, \; H_r < 0, \; H_\beta > 0$$

where $H_q$ is the marginal expected penalty with respect to output, $H_r$ is the marginal expected penalty with respect to changes in quota holding and $H_\beta$ is the marginal expected penalty with respect to changes in the value of the parameter $\beta$. It is assumed that the expected penalty increases with output at any given quota level and decreases in the size of quota holding at any output level. $\beta$ is a shift parameter in the expected penalty function external to the firm. Increases in $\beta$ raise the expected penalty that the firm faces at any given output level and quota violation.

From (2-2-3) the firm's profit function becomes

(2-6-4) $$\pi(q,r) = P.q - c(q) - H(q,r,\beta)$$

The firm chooses output so as to maximize the value of this expression. Differentiating with respect to q gives the first-order condition.

(2-6-5) $\frac{\partial\pi(q,r,\beta)}{\partial q} = P - c'(q) - H_q \leq 0$; < only if $q^* = 0$

Solving (2-6-5) gives the firm's optimal output level as a function of the quota level and the expected penalty function at any given price.

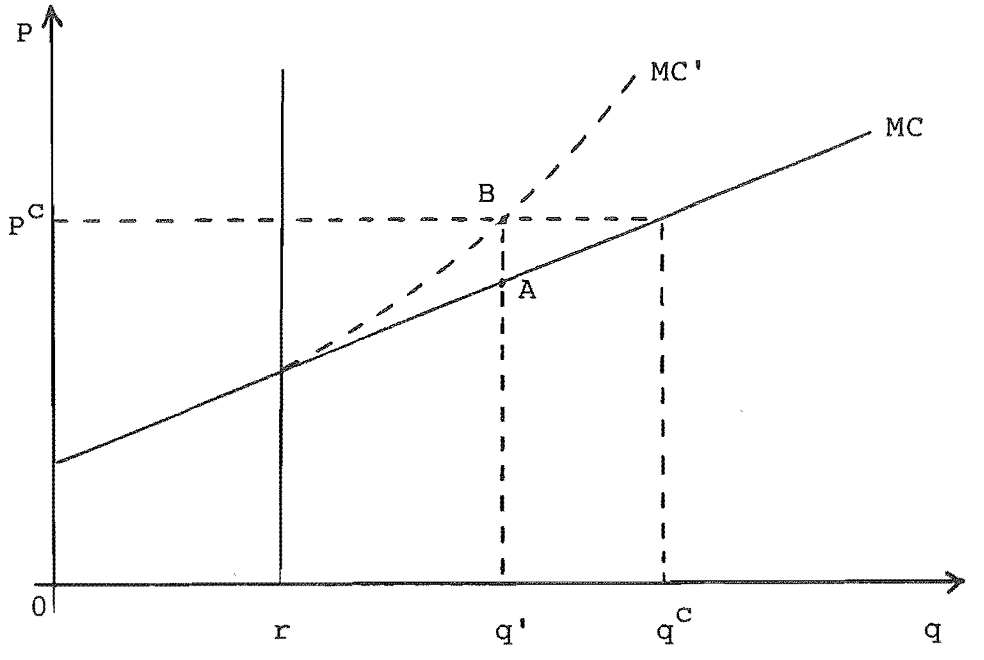
(2-6-6) $q^* = q^*(P,r,\beta)$

Assuming that the market equilibrium price exceeds the marginal production cost for the initial unit of output, and that marginal production cost is non-decreasing in output, the firm produces zero output if and only if it is faced with a zero quota enforced by an expected penalty function such that the marginal expected penalty for the initial unit of production exceeds the difference between marginal production cost and demand price. The most interesting and realistic case, however, is that when prohibited activity occurs and it is this which is treated most extensively in the following analysis.

Equation (2-6-5) shows that, at any positive quantity, the market price must rise by the amount of the marginal expected penalty evaluated at that quantity or alternatively, with increasing marginal costs, quantity will fall at any given price. This is significant when considering the equivalence between price-oriented and quantitative controls. The expected penalty is essentially a tax which must be levied to support the quota. The allocation of quota reduces the quantity of output on which this "tax" is paid.

The firm's marginal cost curve beyond the quota level shifts vertically by the amount of the expected penalty as is shown in Figure 2-6-1 below.

Figure 2-6-1: The effect of an output quota on the individual firm




Recalling that r is the firm's equilibrium holding of quota, the vertical distance between the MC curve and the dashed locus MC' shows the size of the marginal expected penalty at that level of violation such as amount AB at output q' and violation q'-r. With a tradeable quota, the marginal expected penalty represents the minimum additional cost of illegal production or the maximum amount paid to acquire a legal right to produce and is thus the trading price of the quota.

As it has been assumed that an interior solution to equation (2-6-5) exists, the effect of parameter changes on the firm's output decision can be found by totally differentiating the first-order condition. This gives

(2-6-7) $$dP - c''(q)dq - H_{qq}dq - H_{qr}dr - H_{q\beta}d\beta = 0$$

and rearranging

(2-6-8) $$[c''(q) + H_{qq}]dq = dP - H_{qr}dr - H_{q\beta}d\beta$$

PROPOSITION 2-6-1: Suppose that an individual profit maximizing firm in a competitive market is faced with a strictly binding tradeable quota enforced by means of an expected monetary penalty which is dependent on the firm's quota holding and output level and that the firm's profit function is concave;

(i) Supply increases with price.

(ii) An increase in the available quota increases, decreases, or leaves unchanged, the optimal output of the firm at any given price level for which supply is strictly positive if and only if the increase in quota decreases, increases, or leaves unchanged, the marginal expected penalty with respect to output.

(iii) An increase in the expected penalty at any given output and quota level increases, decreases, or leaves unchanged, the output level of the firm and hence increases, decreases, or leaves unchanged the extent of constraint violation, as it decreases, increases, or leaves unchanged, the marginal expected penalty with respect to output.

Proof.

(i) Using (2-6-8) with $d\beta = dr = 0$ gives

(2-6-9) $$\frac{dq^*}{dP} = \frac{1}{c''(q) + H_{qq}} \gtrless 0 \text{ if and only if } c''(q) + H_{qq} \gtrless 0$$

From the second-order conditions, as in Proposition 2-4-1, concavity of the firm's profit function requires overall convexity in output in the firm's penalty-inclusive cost structure. Thus $c''(q) + H_{qq} > 0$ and hence, from (2-6-9), supply increases with price.

(ii) Using (2-6-8) with $d\beta = dP = 0$

(2-6-10) $$\frac{dq^*}{dr} = \frac{-H_{qr}}{c''(q) + H_{qq}} \gtrless 0 \text{ if and only if } H_{qr} \lessgtr 0$$

where $H_{qr}$ is the effect of a change in quota on the marginal expected penalty with respect to output.

(iii) Using (2-6-8) with $dP = dr = 0$

(2-6-11) $$\frac{dq^*}{d\beta} = \frac{-H_{q\beta}}{c''(q) + H_{qq}} \gtrless 0 \text{ as } H_{q\beta} \lessgtr 0$$

where $H_{q\beta}$ is the effect of a change in the expected penalty on the marginal expected penalty with respect to output at any given output level and quota holding.

□

Result (i) is intuitively appealing. With a convex penalty-inclusive cost function, profitable expansion in output requires an increasing market price. If the expected penalty function is convex in output, in addition to the normal condition on production cost, supply is less price responsive in the presence of the quota than in the unregulated environment.

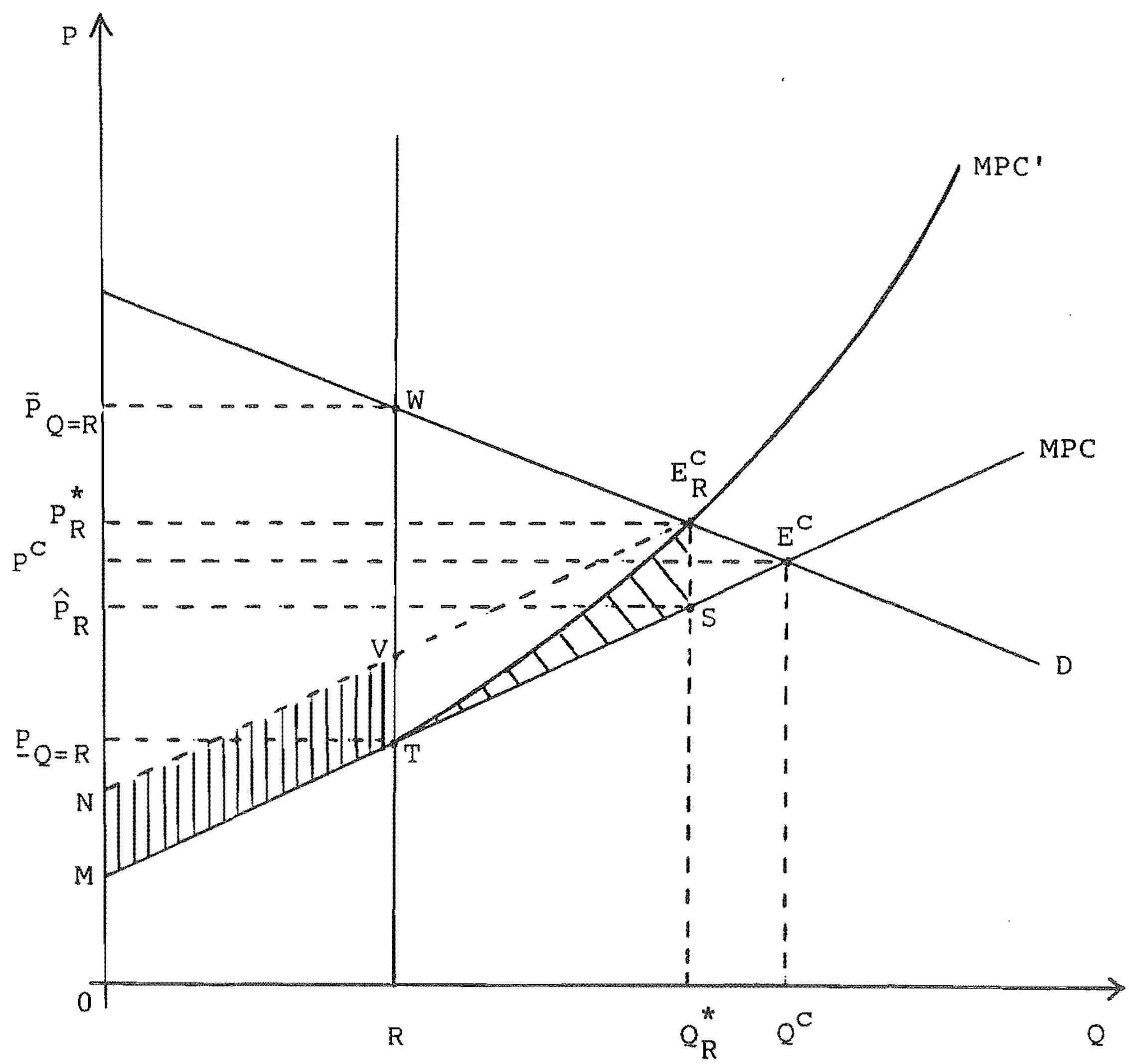
Result (ii) raises the possibility of the perverse result that when the quota level is increased or decreased, equilibrium output changes in the opposite direction. Thus for instance a tightening of the quota on the firm could lead to an expansion in the firm's output at any price. This represents a rightward shift in its supply curve. Given the other assumptions of the model, a necessary and sufficient condition to preclude this occurrence is $H_{qr} < 0$. This result together with the price responsiveness result from part (i) of the Proposition will be discussed further in Section 2-7 when specific forms of the expected penalty function and their behavioural implications are considered.

Finally, result (iii) also contains the possibility of a perverse outcome. As outlined in Section 2-4, the generally accepted result in the deterrence literature is that increased penalties reduce non-compliance. From (2-6-11) this occurs if and only if $H_{q\beta} > 0$. That is, the increase in expected penalty must of necessity raise the marginal expected penalty with respect to output in order for increased penalties to facilitate compliance. As argued with reference to (2-6-5), $H_q$ is essentially the price of illegal activity. The assumption that $H_{q\beta} > 0$ is nothing more than a requirement that the increase in the expected penalty structure raises the price of illegal activity. A well-designed enforcement strategy could be expected to exhibit this property but the alternative policy is further discussed in Section 2-7.

This result showing that increased penalties increase compliance with regulations is consistent with Proposition 2-6-2 and the same comments apply.

The first-order condition given in (2-6-5) showed that, with a non-zero marginal expected penalty, the firm's marginal costs, inclusive of the expected penalty, increase with the imposition of a quota leading to the leftward or upward shift in the supply curve shown in Figure 2-6-1. The industry result is illustrated in Figure 2-6-2 below.

Figure 2-6-2. The effect of an output quota on the industry




Curves MPC and MPC' are constructed on the assumption that there is technical and economic efficiency within the industry so that with identical firms, they differ only in aggregation scale from those of the individual firm.

Beyond the level of the aggregate quota R, the industry supply curve shifts vertically by the marginal expected penalty at each quantity and consequent violation level.[8]

As shown in Figure 2-6-2, the competitive regulated industry equilibrium occurs at some point such as $E_R^C$ where the marginal expected penalty is equal to the difference between the demand price and private marginal production cost evaluated at that quantity.

Thus at the regulated competitive equilibrium $E_R^C$

$$(2\text{-}6\text{-}12) \qquad H_q(q_r^*, r, \beta) = H_Q(Q_R^*, R, \beta) = P_R^* - \hat{P}_R$$

where $P_R^* - \hat{P}_R$ is the difference between the demand price and private marginal production cost at the regulated equilibrium quantity $Q_R^*$. The assumption of identical competitive firms ensures that, at the regulated equilibrium, the marginal expected penalty with respect to the expansion of industry output, $H_Q$, is identical to that faced by any individual firm in the industry.

The regulated competitive equilibrium at $Q_R^*$ is such that

$$(2\text{-}6\text{-}13) \qquad R \leq Q_R^* \leq Q^C \ ; \ Q_R^* = zq_r^*$$

Expression (2-6-13) shows that there are three possibilities for the regulated competitive equilibrium. Firstly, it may coincide with the unregulated competitive equilibrium. With a non-negative expected penalty, as given in expression (2-6-3), the quota sets a lower bound on industry output while the unregulated competitive equilibrium sets an upper bound. This result then requires either that the quota be set at $Q^C$ in which case the expected penalty is irrelevant to the decisions of the firm or that the expected penalty is zero in which case the specific quota level is irrelevant.

Secondly, the regulated competitive equilibrium may coincide with an output quota which is strictly less than the unregulated competitive equilibrium. This requires that the marginal expected penalty be at least as great as the difference between the demand price and the marginal production cost at the quota level. Thus in Figure 2-6-2,

$$(2\text{-}6\text{-}14)\quad \lim_{q\to r^+} H_q(q,r,\beta) = \lim_{Q\to R^+} H_Q(Q,R,\beta) \geq \bar{P}_{Q=R} - \underline{P}_{Q=R}$$

where $\bar{P}_{Q=R} - \underline{P}_{Q=R}$ is the difference between the demand price and private marginal production cost at the quota level R, would ensure that the regulated competitive equilibrium occurs at W with zero violation of the quota.

Finally, the regulated competitive equilibrium may occur at some intermediate point between $E^C$ and W as is shown in Figure 2-6-2. This will be the case if the marginal expected penalty at the quota level R is less than the value of the difference between the demand price and private marginal cost at the quota level R and is that treated most extensively in the following analysis.

From Figure 2-6-2 it appears that the regulated competitive equilibrium involves an increased market price and reduced market quantity when compared with the unregulated competitive equilibrium. This leads to the following proposition.

PROPOSITION 2-6-2: Suppose that an industry comprising a finite number of identical competitive firms faces a market demand curve with non-zero finite elasticity and is subject to a strictly binding tradeable output quota enforced by means of an expected monetary penalty function which is dependent on the individual firm's quota holding and output level;

(i) Assuming that an increase in available quota reduces the marginal expected penalty with respect to output at any given output level and quota violation, an increase in quota increases industry equilibrium output and reduces equilibrium price.

(ii) Assuming that an increase in the expected penalty function raises the marginal expected penalty with respect to output at any given output level and quota holding, an increase in the expected penalty reduces industry equilibrium output and raises the equilibrium price.

Proof.

(i) From (2-6-6), (2-2-2) and (2-2-7), it is evident that the market equilibrium price is a function of the expected penalty thus at the industry equilibrium

$$D(P(R,\beta)) = S(P(R,\beta),R,\beta) \tag{2-6-15}$$

where $S(P(R,\beta),R,\beta)$ is industry supply within the regulated environment defined as the summation of individual regulated supply curves which are derived from (2-6-5).

Totally differentiating (2-6-15) gives

$$(2\text{-}6\text{-}16) \quad \frac{\partial S(.)}{\partial P}\frac{\partial P(R,\beta)}{\partial R} + \frac{\partial S(.)}{\partial R} + \frac{\partial S(.)}{\partial P}\frac{\partial P(R,\beta)}{\partial \beta} + \frac{\partial S(.)}{\partial \beta}$$

$$- D'(P(R,\beta))\frac{\partial P(R,\beta)}{\partial R} - D'(P(R,\beta))\frac{\partial P(R,\beta)}{\partial \beta} = 0$$

Rearranging with $d\beta = 0$

$$(2\text{-}6\text{-}17) \quad \frac{\partial P^*(R,\beta)}{\partial R} = \frac{\partial S(.)/\partial R}{D'(P(R,\beta)) - \partial S(.)/\partial P} < 0$$

given the assumption that $H_{qr} < 0$ which from Proposition 2-6-1 and (2-2-7) ensures that $\partial S(.)/\partial R > 0$. Noting from (2-6-15) that

$$(2\text{-}6\text{-}18) \quad Q_R^* = D(P^*(R,\beta))$$

where $Q_R^*$ is the regulated industry equilibrium output level,

$$(2\text{-}6\text{-}19) \quad \frac{\partial Q_R^*}{\partial R} = D'(P(R,\beta))\frac{\partial P^*(R,\beta)}{\partial R} > 0$$

using (2-6-17)

(ii) Rearranging (2-6-16) with $dR = 0$

$$(2\text{-}6\text{-}20) \quad \frac{\partial P^*(R,\beta)}{\partial \beta} = \frac{\partial S(.)/\partial \beta}{D'(P(R,\beta)) - \partial S(.)/\partial P} > 0$$

given the assumptions that $H_\beta > 0$ and $H_{q\beta} > 0$ which from Proposition 2-6-1 and (2-2-7) ensure that $\partial S(.)/\partial \beta < 0$. From (2-6-18)

$$(2\text{-}6\text{-}21) \quad \frac{\partial Q_R^*}{\partial \beta} = D'(P(R,\beta))\frac{\partial P^*(R,\beta)}{\partial \beta} < 0.$$

□

Equation (2-6-17) shows that an increase in the available output quota,which increases supply at every price, decreases the market equilibrium price of the commodity. Multiplying this result by the price responsiveness of the demand curve in (2-6-19) gives the effect of the change in quota on the equilibrium quantity in the market. Given the new market equilibrium price, the individual firm's regulated output can be found by solving (2-6-5) at this price which gives $q_r^*$ consistent with (2-6-13). Equations (2-6-20) and (2-6-21) provide a similar analysis of the effects of a change in the expected penalty on market equilibrium output and price levels.

From Proposition 2-6-2 and the discussion of the conditions that generated each of the possible outcomes in expression (2-6-13) it is evident that, with an appropriate mix of quota and expected penalty, it is possible to restrict industry output to the socially optimal level $\hat{Q}$ in Figure 2-2-2. This requires that either,

$$(2\text{-}6\text{-}22) \qquad R = \hat{Q} \ ; \ \lim_{Q \to R^+} H_Q(Q, \hat{Q}, \beta) \geq \overline{P}_{\hat{Q}} - \underline{P}_{\hat{Q}}$$

or

$$(2\text{-}6\text{-}23) \qquad R < \hat{Q} \ ; \ H_Q(\hat{Q}, R, \beta) = \overline{P}_{\hat{Q}} - \underline{P}_{\hat{Q}}.$$

Expression (2-6-22) describes the situation of a quota set at the desired output level accompanied by an expected penalty which prohibits violation whereas expression (2-6-23) uses a more severely restrictive quota with an expected penalty structure which produces violation consistent with the desired output target.

Whether or not this occurs and which method is chosen depends on the objectives of the regulator as will be discussed in Section 2-8. In the situation illustrated in Figure 2-6-2, the area $TE_R^CS$ shows the total expected fine payments incurred by the industry while the vertical distance $SE_R^C$ shows the marginal expected penalty at $Q_R^*$. From the discussion of Figure 2-6-1 this becomes the equilibrium trading price of the quota and determines the aggregate rent which accrues to the industry shown by area MNVT which equals $(P_R^* - \hat{P}_R)R$. The relative sizes of these magnitudes will also prove significant when the objectives of the regulator are discussed.

## 2-7 FORMS OF THE EXPECTED PENALTY FUNCTION

In Sections 2-4, 2-5 and 2-6, the properties of expected penalty functions necessary and/or sufficient to provide certain results were established and these results were qualitatively illustrated for the individual firm and the total industry. The analysis however avoided consideration of specific functional forms which may or may not exhibit the assumed properties. Several types of function are now examined.

Firstly, the expected penalty may consist of a lump-sum fine incurred with constant probability regardless of the size of the violation. In this case there is a known and fixed expected cost of illegal activity. Once the non-compliance decision has been made, the marginal expected penalty is zero and the per unit expected penalty cost declines as the extent of violation increases. This type of penalty structure is biased against relatively small violations of the constraint so that any evasion that did occur would be extensive. The firm's profit function becomes discontinuous and hence the first-order conditions do not apply.

PROPOSITION 2-7-1: Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with non-zero finite elasticity and is subject to either a strictly positive unit rate sales tax or strictly binding tradeable output quota. If the regulatory constraint is enforced by means of a lump-sum expected monetary penalty which is independent of the extent of illegal behaviour, then for sufficiently low values of the penalty firm and industry behaviour is unaffected while for sufficiently high values of the penalty full compliance is ensured.

Proof.

At the unregulated competitive industry equilibrium, each individual firm is maximizing profits subject to the technological and financial constraints of its production process. At the market equilibrium price, its output is given by the value of its supply function at that price. From the assumption of a fixed number of identical competitive firms, each firm in the industry produces the same quantity of output and earns the same amount of profit. This level of profit is now denoted by $\pi_{CE}$.

As has been previously argued, the imposition of any additional binding constraint on the firm will of necessity reduce its profit, ceteris paribus, below the competitive equilibrium level $\pi_{CE}$. The level of profit which the firm would enjoy under full compliance with the regulatory constraint is denoted as $\pi_{FC}$. Hence,

(2-7-1) $\quad 0 \leq \pi_{FC} < \pi_{CE}$

At the prevailing equilibrium price, profit associated with full compliance is strictly less than that at the profit-maximizing output level. Profits cannot be negative under full compliance with the regulatory constraint as the firm always has the option of producing nothing which, with no fixed production costs, generates zero profit.

If the regulation is not enforced then profit maximization dictates that the regulated equilibrium coincides with the unregulated competitive equilibrium. If, as assumed here, the regulation is enforced by a lumpsum expected penalty, there is a fixed and known cost of illegal behaviour. This is equivalent to a poll tax or pure profit tax which has no marginal effects. First-order conditions therefore are not applicable.

Denoting the size of the lumpsum expected penalty by K, profit at the unregulated competitive equilibrium output level within the regulated environment, defined as $\pi_{CEP}$, is

(2-7-2) $\quad \pi_{CEP} = \pi_{CE} - K \; ; \quad K > 0$

For all values of K such that

(2-7-3) $\quad \pi_{CE} - K > \pi_{FC}$

individual firm behaviour is unaffected by the regulation and enforcement. Profitability is reduced but marginal conditions are unchanged and therefore the regulated industry equilibrium coincides with the unregulated competitive industry equilibrium.

For all values of K such that

(2-7-4) $\quad \pi_{CE} - K < \pi_{FC}$

individual profit maximization dictates that full compliance is the optimal strategy. The regulated industry equilibrium then exhibits full compliance with the regulatory constraint.

Denoting by $\bar{K}$ the value of the lumpsum expected penalty such that

(2-7-5) $\quad \bar{K} = \pi_{CE} - \pi_{FC}$

from (2-7-3) and (2-7-4), for all $K < \bar{K}$ industry behaviour is unaffected by the introduction of the regulation while for all $K > \bar{K}$, full compliance is ensured.

□

A second possibility is that of a constant unit rate expected penalty set at some level k. In this case for a tax $G_q = -G_x = k$ and $G_{qq} = G_{xx} = G_{xq} = 0$, and for a quota $H_q = -H_r = k$ and $H_{qq} = H_{qr} = 0$ given fixed values of the parameters $\alpha$ and $\beta$. The slope of the firm's supply curve in both cases is unchanged by this form of expected penalty function but it is possible to ensure full compliance by setting the rate sufficiently high.

<u>PROPOSITION 2-7-2</u>: <u>Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with non-zero finite elasticity and is subject to either a strictly positive unit rate sales tax or strictly binding tradeable output quota enforced by means of a constant unit rate expected monetary penalty.</u>

(i) <u>For sufficiently low values of the expected penalty rate, no output is declared in the sales tax case.</u>

(ii) <u>In the case of an output quota, changes in the level of quota, which remain binding at any given expected penalty rate, do not affect the regulated equilibrium.</u>

(iii) <u>In each case, full compliance can be ensured by sufficiently high expected penalty rates.</u>

<u>Proof</u>.

(i) Using (2-4-2) and (2-4-13) profit from selling an additional unit of declared and undeclared output, denoted by $\pi'_D$ and $\pi'_U$ respectively, compares as follows.

(2-7-6) $$\pi'_D = P - c'(q) - t \gtrless P - c'(q) - k = \pi'_U \,;\; t > 0,\; k \geq 0;$$

if and only if

(2-7-7) $$k \gtrless t$$

Profit maximization ensures that the declaration strategy on the marginal unit is determined by the action which increases profit by the greatest amount. For the unit to be produced at all, at least one of $\pi'_D$, $\pi'_U$ must be non-negative.

If $k = 0$, which corresponds to the situation of no enforcement, then clearly $\pi'_U > \pi'_D$ and, assuming that $\pi'_U \geq 0$, the output unit will be produced and not declared. By the same argument, (2-7-6) and (2-7-7) show that for any expected penalty rate which is less than the tax rate, the marginal unit of output, if produced, will be undeclared.

As the marginal expected penalty and tax rate are constant, any output that is produced will be undeclared if $k < t$.

(ii) With a constant unit rate expected penalty, a change in the quota level which remains binding on behaviour does not alter the marginal expected penalty with respect to output. Hence $H_{qr} = 0$ at any given value of the penalty parameter and from (2-6-10) and (2-2-7) individual behaviour and the regulated industry equilibrium are unaffected by the change in quota.

(iii) In the case of a sales tax, from (2-7-6) and (2-7-7) any output that is produced will be fully declared if $k > t$. In the case of an output quota, profit on the marginal unit of output using (2-6-5) is

(2-7-8) $$\pi_r' = P - c'(q) - k$$

From profit maximization, the marginal unit of output will be produced if and only if $\pi_r' \geq 0$. Using (2-6-14), full compliance with an output quota is ensured if marginal profit is zero at the level of the quota. For the industry and the firm, this requires an expected penalty rate $\bar{k}$

(2-7-9) $$\bar{k} = P^*(R) - c'(R)$$

where $P^*(R)$ is the market equilibrium price at the quota level of output, being the demand price at this output, and $c'(R)$ is the level of marginal production cost at the quota level of output for the industry.

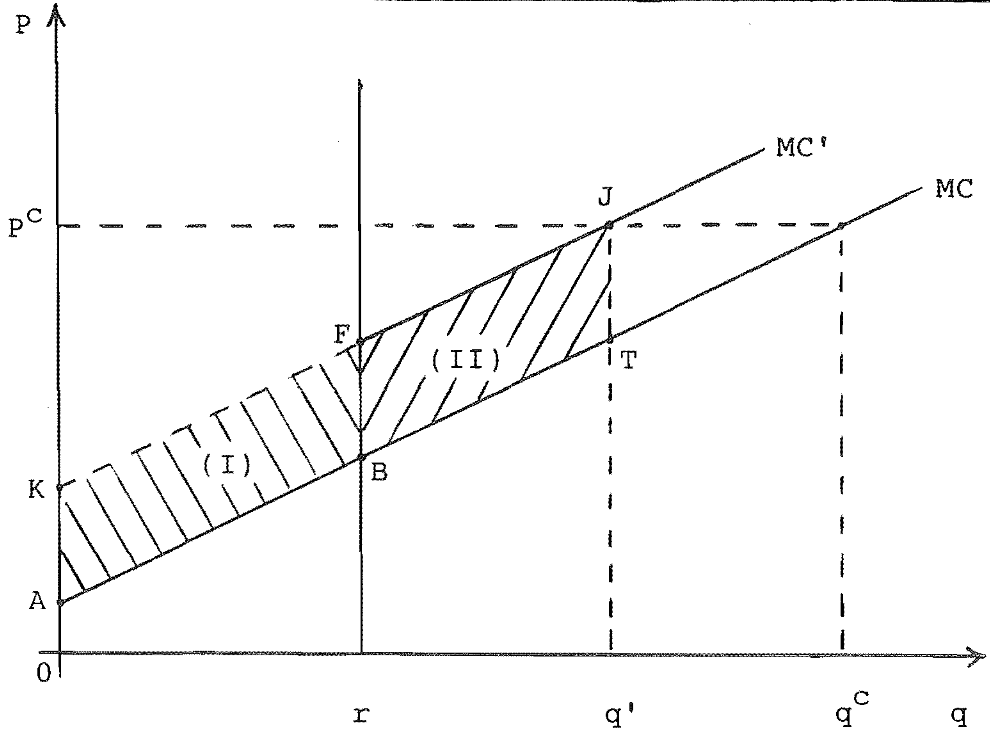
From the assumption of identical firms, this is equal to $c'(r)$ for the individual firm where $r = R/z$ and $z$ is the number of firms in the industry.

At any expected penalty rate $k > \bar{k}$ therefore, full compliance with the output quota is assured.

□

The situation of an output quota enforced by a constant unit rate expected penalty in accordance with Proposition 2-7-2 is illustrated in Figure 2-7-1.

Figure 2-7-1: The effect of an output quota with a constant unit rate expected penalty




With the constant unit rate expected penalty, (2-6-9) shows that the slope of the firm's supply curve is identical in both regulated and unregulated environments. Beyond the quota level r, the firm's regulated supply curve shifts vertically by the amount of the expected penalty rate and runs parallel with the original curve. This is shown by the stepped marginal cost curve ABFJ. All output in excess of the quota attracts the same expected penalty shown by FB in the diagram. As was established in Section 2-6, arbitrage dictates that this also becomes the unit trading price of the quota. The firm's optimal regulated output at price $P^C$ is then given from the new supply curve MC' at q'.

Once a quota $r \leq q'$ is established together with the expected unit rate penalty FB = KA, the size of the aggregate shaded area AKJT, which represents the increase in costs on 0q' units of output produced in the regulated environment, is uniquely determined. With output invariant, changes in the quota level merely serve to determine the decomposition of the aggregate magnitude between area (I), the rent accruing to the original holder of the quota, and area (II) representing the total expected penalty payment.

Thirdly the expected penalty function may exhibit changing marginal rates which are either increasing or decreasing with the extent of undeclared output. Figure 2-7-2 illustrates the case of an expected penalty function for which the marginal expected penalty increases with tax evasion.

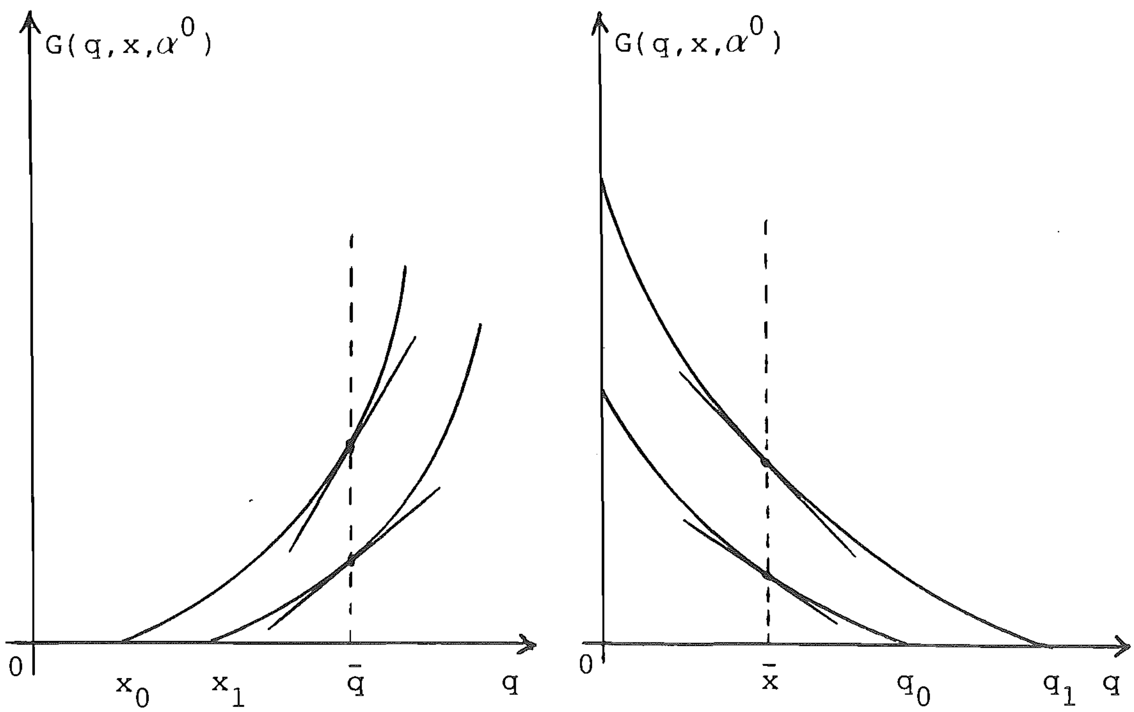
Panel (i) illustrates the expected penalty function in output space for two given values of declared output while panel (ii) shows the same function in declared output space for two values of actual production. This function is consistent with the results shown in Figures 2-4-1 and 2-5-1. The increase in declaration in panel (i) decreases the marginal expected penalty with respect to output at a given output level while an increase in actual output increases the negative slope in panel (ii). Hence $G_{xq} = G_{qx} < 0$

and the function satisfies the conditions of Propositions 2-4-1 and 2-5-1 (i) required to generate the intuitive results presented there.

Figure 2-7-2: An expected penalty function that increases with tax evasion

(i) Output space with a given level of declared output

(ii) Declaration space with a given level of actual output

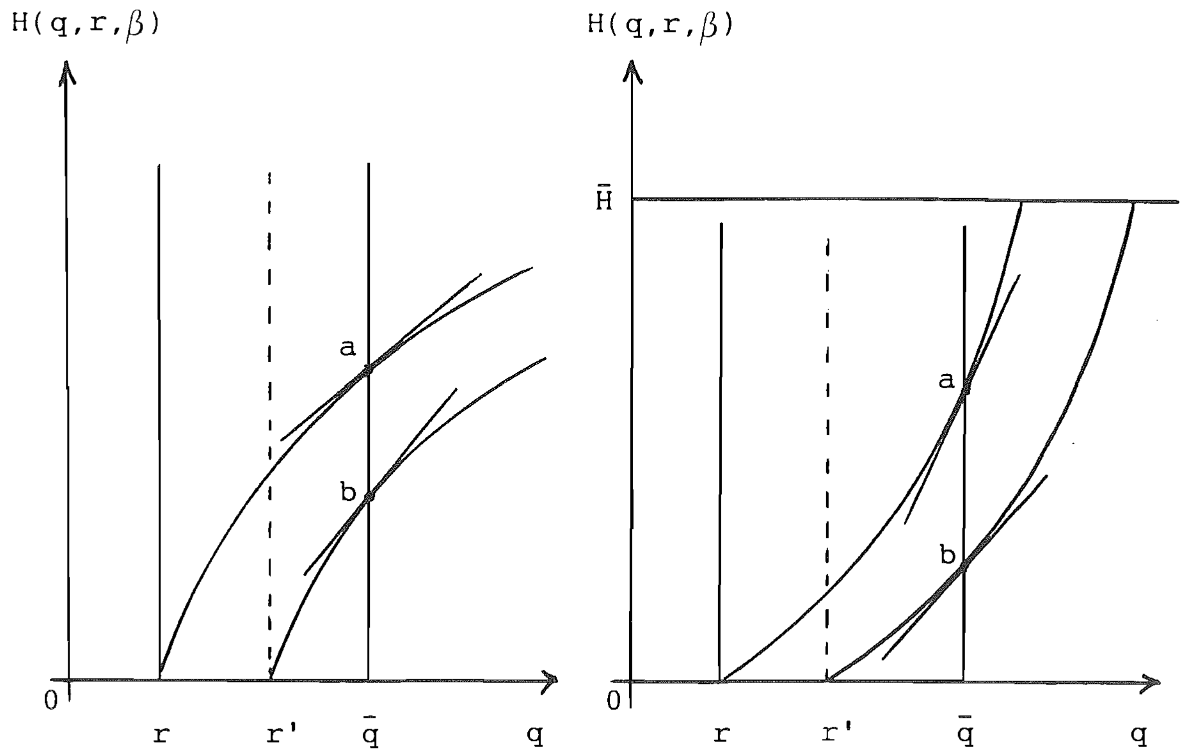



Not all enforcement regimes, however, exhibit the properties which ensure that the firm's penalty-inclusive profit function is strictly concave. The expected penalty function may be only locally convex or locally and globally concave. Two such examples are illustrated in Figure 2-7-3 below for the case of an output quota. Each curve is drawn for a fixed value of the parameter $\beta = \beta_0$.

Figure 2-7-3: Expected penalty functions that are non-convex in the extent of constraint violation

(i) Concave expected penalty function

(ii) Bounded expected penalty function




Panel (i) illustrates an expected penalty function that is concave in output so that $H_{qq} < 0$. The effect of an increase in quota from r to r' causes the value of the marginal expected penalty, evaluated at any given output level, to rise. Thus the slope at 'b' exceeds that at 'a' and $H_{qr} > 0$.

Panel (ii) portrays an expected penalty function that is locally convex in output so that for particular output levels $H_{qq} > 0$. The effect of an increase in quota from r to r' causes the value of the marginal expected penalty, evaluated at these output levels, to fall. Thus the slope at 'b' is less than that at 'a' and $H_{qr} < 0$. The expected penalty, however, is bounded from above by some finite value $\bar{H}$. For all output levels beyond that at which the expected penalty reaches its maximum value, given the available quota level, the marginal expected penalty with respect to output is zero.

PROPOSITION 2-7-3: Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with non-zero finite elasticity and is subject to a strictly binding output quota enforced by means of an expected monetary penalty that is concave in output. Assuming that the convexity of production cost dominates the non-convexity of the expected penalty and that an increase in available quota increases the marginal expected penalty with respect to output at any output level;

(i) Supply increases with price.

(ii) An increase (decrease) in available quota decreases (increases) supply at every price and hence raises (lowers) market equilibrium price and reduces (increases) industry equilibrium output.

(iii) <u>An increase in the expected penalty reduces supply at every price and hence raises market equilibrium price and lowers industry equilibrium output if it acts to increase the marginal expected penalty with respect to output at any output level which exceeds the quota</u>.

<u>Proof</u>.

(i) Given the assumption that the convexity of production cost in output dominates the non-convexity of the expected penalty function, $c''(q) + H_{qq} > 0$ and, from (2-6-9), individual firm supply increases with price in the regulated environment. Hence, by the aggregation in (2-2-7), the industry supply curve is upward sloping.

(ii) With $H_{qr} > 0$ and $c''(q) + H_{qq} > 0$, then, from (2-6-10), an increase (decrease) in available quota, such that the quota remains binding on behaviour, will reduce (increase) individual supply at every price. By aggregation, industry supply will also fall (rise) at every price and hence, from (2-6-17) and (2-6-19), industry equilibrium output will decrease (increase) while market equilibrium price rises (falls) as a result of the increase (decrease) in quota.

(iii) If $H_{q\beta} > 0$, from (2-6-11) an increase in the expected penalty reduces individual supply at every price given that $c''(q) + H_{qq} > 0$. By aggregation, industry supply also declines at every price and hence, from (2-6-20) and (2-6-21), industry equilibrium output is reduced and market equilibrium price raised by the increased expected penalty associated with quota violations.

□

The situation of an industry subject to a concave expected penalty function in accordance with Proposition 2-7-3 is illustrated in Figure 2-7-4 below.

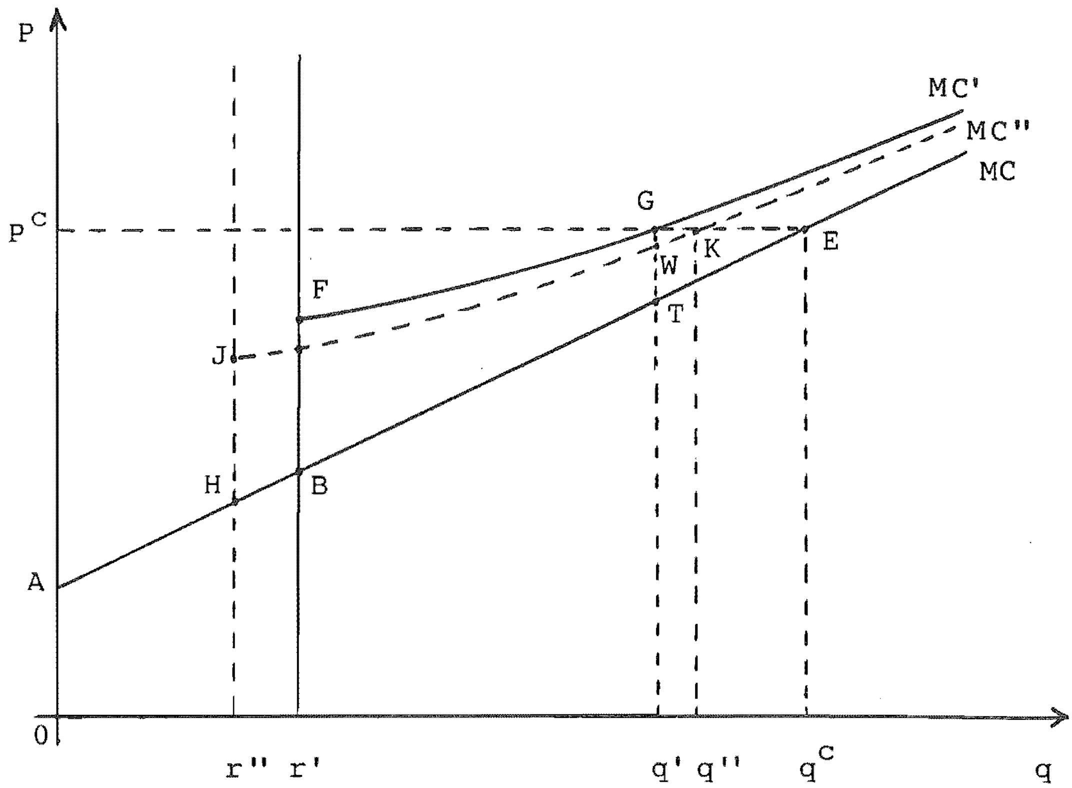
Panel (i) illustrates results (i) and (ii) of Proposition 2-7-3 for the individual firm. The supply curve in the unregulated environment is shown by curve MC. Each of MC' and MC" is drawn for a fixed value of the penalty parameter $\beta = \beta_0$.

From (2-6-5) the supply curve of the firm shifts vertically by the amount of the marginal expected penalty with respect to output for all output levels exceeding the quota. As illustrated in panel (i) of Figure 2-7-3 the marginal expected penalty with respect to output for a concave expected penalty function is maximized for the initial unit of illegal output and declines thereafter. The assumption of a non-negative marginal expected penalty ensures that the regulated supply curve never lies below the unregulated curve.

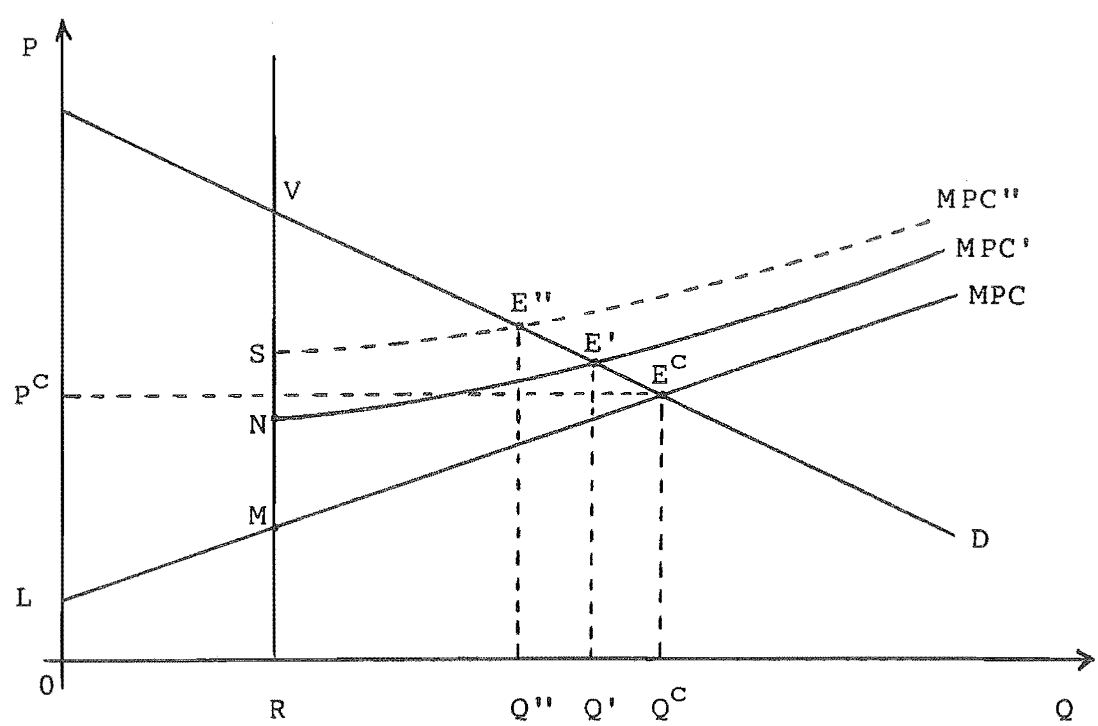
At the initial quota level r' the penalty-inclusive supply curve is given by the line ABFG denoted as MC'. Optimal output at the original unregulated market equilibrium price $P^c$ is given by q' at point G. A reduction in the quota from r' to r", by the assumption that $H_{qr} > 0$, serves to reduce the marginal expected penalty at any output level which exceeds the original quota level. Thus at q' the marginal expected penalty with respect to output falls from

Figure 2-7-4: The effect of an output quota enforced by a concave expected penalty function

(i) The effect on an individual firm of a change in available quota




(ii) The effect on the industry of an increase in the expected penalty structure



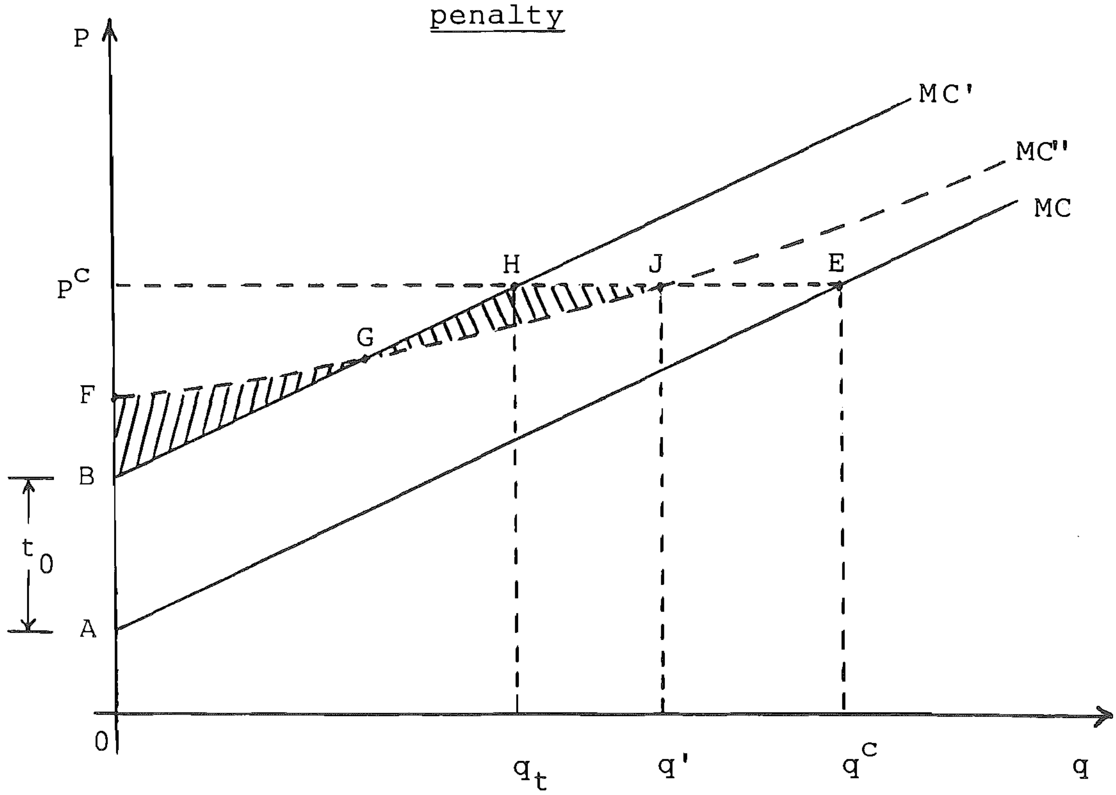
TG to TW. The effect of the reduction in quota is then to increase the optimal output level of the firm at $P^C$ from q' to q" associated with point K. The reduction in quota therefore induces an increase in supply at any price level for which illegal output is produced. This is a perverse outcome which severely limits the effectiveness of a regulatory body in restraining the output of an industry.

Panel (ii) of Figure 2-7-4 illustrates the effect of an increase in expected penalty on the industry equilibrium. The unregulated industry supply curve is given by MPC. Curves MPC' and MPC" denote regulated supply at penalty levels $\beta'$ and $\beta''$ respectively with $\beta'' > \beta'$. Following the assumption that $H_{q\beta} > 0$, the marginal expected penalty with respect to output at any illegal output level is raised by the increase in the expected penalty structure. Thus the marginal expected penalty for the initial illegal unit of output rises from MN to MS. As a result, the increase in the penalty parameter from $\beta'$ to $\beta''$ reduces industry equilibrium output from Q' at point E' to Q" associated with E". Full compliance with the quota can be ensured by increasing the structure of the expected penalty until the marginal expected penalty for the initial unit of illegal output is equal to VM.

The effect of such an expected penalty function on an industry regulated by a unit rate sales tax can be examined with reference to Figure 2-7-5 below.

As in the above case of the quota, the convexity of production cost is assumed to dominate the concavity in output of the expected penalty function and hence $c''(q) + G_{qq} > 0$. From the assumed concavity of the expected penalty function $G_{xx} < 0$, which, using (2-4-20) and (2-4-21), shows that the firm's supply curve in the regulated environment remains upward sloping.

Figure 2-7-5: The effect of a unit rate sales tax enforced by a concave expected monetary penalty




The unregulated supply curve is shown on the diagram by MC and the tax-inclusive supply curve at rate $t_0$ is given by MC'. The curve labelled MC" illustrates the firm's penalty-inclusive marginal cost curve given that no output is declared. As with the quota, the marginal expected penalty with respect to output is maximized for the initial illegal unit of output and declines thereafter. The assumption that $G_q \geq 0$ again ensures that MC" does not lie below the unregulated curve MC.

Clearly if the marginal expected penalty for the initial illegal unit is less than the tax rate, from (2-4-14) zero declaration is the optimal strategy. When, as is the case in Figure 2-7-5, $H_q(0,0,\alpha^0) > t_0$, the optimal declaration strategy at any price such as $P^c$ is determined by considering the profit level at the output levels q' and $q_t$ associated with zero declaration and full declaration respectively. Any intermediate strategy is not optimal in these circumstances as, given that $G_{qx} > 0$, reducing the level of declared output increases the marginal return at every level of illegally produced output.

Examining Figure 2-7-5, the comparison of profit levels at q' and $q_t$ involves consideration of the relative sizes of areas FGB and GHJ. Beginning at point H which is associated with the fully declared output level $q_t$, a move to zero declaration at J reduces profit on initial units of output by FGB and raises it by GHJ on later output units. At any price level then, output is fully declared if FGB > GHJ and no output is declared if GHJ > FGB.

Assuming initially that output is fully declared, an increase in the tax rate reduces area FGB and raises GHJ. Sufficiently high tax rates result in no output being declared. Successive increases in the tax rate from low levels, at any given expected penalty structure, result in reductions in the regulated equilibrium output level of the industry until some critical rate is reached when regulated equilibrium output jumps suddenly to that level associated with zero declaration.

Conversely, assuming that initially no output is declared, optimal output given price $P^c$ is q'. Increases in the structure of the expected penalty function, given that $G_{q\alpha} > 0$, increase the size of area FGB and reduce GHJ while lowering the regulated equilibrium output level of the firm and industry. Sufficiently high levels of the expected penalty function result in output being fully declared. Successive increases in the structure of the expected penalty function, which raise the marginal expected penalty with respect to output at any illegal output level, result in reductions in the regulated equilibrium output level of the industry until some critical penalty structure is reached when regulated equilibrium output jumps to that level associated with full declaration.

These results are summarized in the following Proposition.

PROPOSITION 2-7-4: Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with non-zero finite elasticity and is subject to a strictly positive unit rate sales tax enforced by means of a concave expected monetary penalty. Assuming that the convexity of production cost dominates the concavity in output of the expected penalty function, supply increases with price. Optimal firm and industry behaviour in the regulated environment is determined by global profit maximization which may or may not produce outcomes consistent with first-order conditions. In particular it is possible that changes in the rate of sales tax given any expected penalty structure, or alternatively, in the expected penalty structure at any given tax rate, induce discontinuous responses in regulated equilibrium output levels.

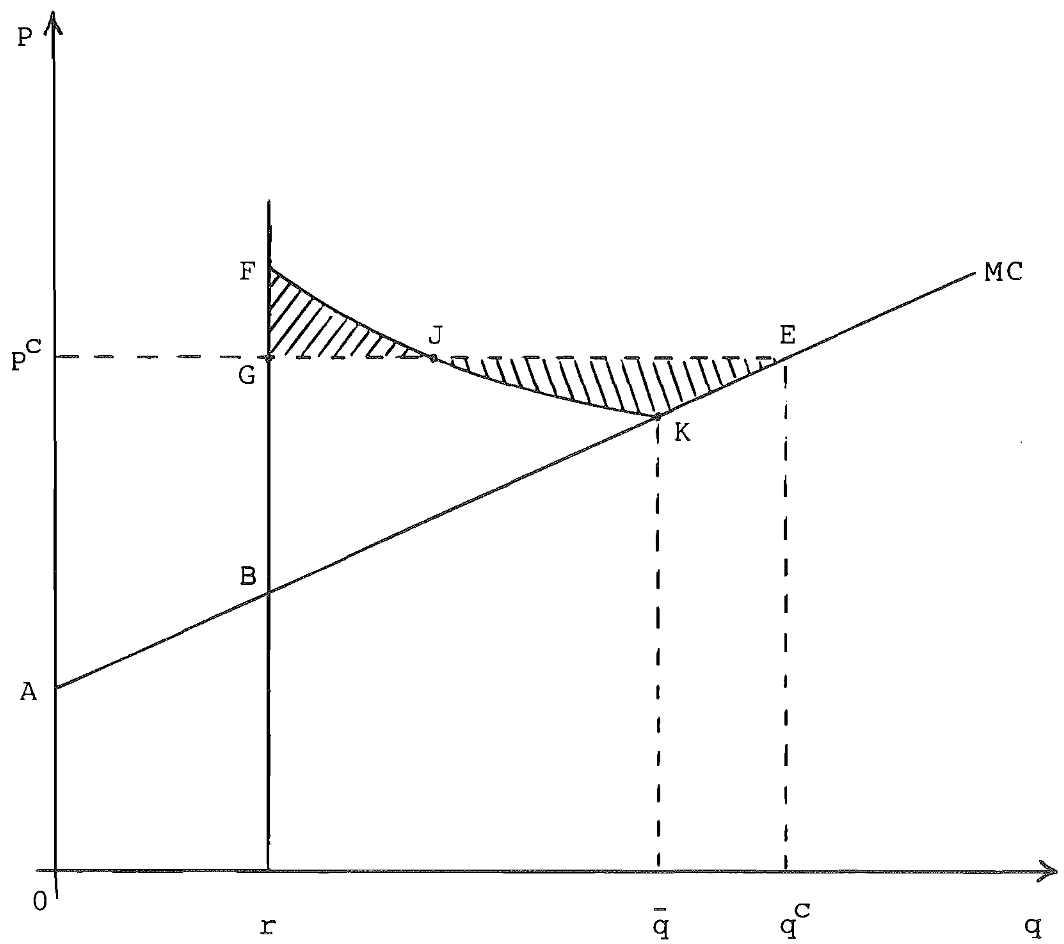
It is possible with an expected penalty function that is concave in output that this concavity outweighs the convexity of unregulated production cost so that, in the case of a quota, $c''(q) + H_{qq} < 0$. In this case, from (2-6-9), the firm's penalty-inclusive supply function is negatively sloped for at least some range of output levels. Given the assumptions of strictly convex production cost and that the marginal expected penalty with respect to output is non-negative, the regulated supply curve cannot be negatively sloped for all illegal output levels.

It is likely in these circumstances that the expected penalty is bounded as is the case in Figure 2-7-6 below.

The concavity of the expected penalty function ensures that the marginal expected penalty with respect to output is maximized at the initial illegal unit of output. The marginal expected penalty with respect to output declines as illegal output increases until output level $\bar{q}$ where it becomes zero. At this level, the expected penalty is maximized at size BFK.

Optimal output at any price is found at that level which maximizes global profits. In this situation, the first-order conditions are irrelevant and optimal output occurs at either the quota level r or the unregulated competitive equilibrium level $q^c$. This involves a consideration of the relative sizes of GFJ and JEK.

Figure 2-7-6: The effect of an output quota enforced by a bounded concave expected penalty function




The penalty-inclusive marginal cost curve is shown by ABFKE. Beginning at the full compliance output level r, a decision to produce at $q^c$ reduces profit by GFJ but increases it by JEK. Optimal output therefore occurs at full compliance with the quota when GFJ > JEK and at the unregulated equilibrium level $q^c$ where JEK > GFJ.

Following panel (i) of Figure 2-7-3, $H_{qr} > 0$ and a reduction in available quota reduces the expected penalty at any given illegal output level thus increasing area JEK without increasing GFJ or the total size of the expected penalty. Sufficiently small levels of quota at any given expected penalty structure will result in optimal output at the unregulated competitive equilibrium output level $q^c$.

Conversely, increases in the expected penalty structure at any given quota level increase area GFJ and decrease JEK. This reduces the likelihood that the unregulated competitive equilibrium output level is optimal and indeed, for sufficiently high levels of the expected penalty function, full compliance is the optimal strategy.

Similar results occur when the industry is regulated by a unit rate sales tax. In each case changes in the regulatory constraint or in the structure of the expected penalty function may at some critical level cause the regulated equilibrium output level of the firm and industry to switch from the unregulated competitive equilibrium value to that associated with full compliance with the constraint.

These results are summarized in the following Proposition.

PROPOSITION 2-7-5: Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with nonzero finite elasticity and is subject to either a strictly positive unit rate sales tax or a strictly binding tradeable output quota each enforced by an expected monetary penalty.

Assuming that the expected penalty function is concave and bounded, the concavity of the expected penalty dominates the convexity of production cost so that the firm's regulated marginal cost curve is negatively sloped over a certain output range. Optimal output is determined by global profit maximization and occurs at either the unregulated competitive equilibrium level or at that level associated with full compliance with the constraint. Changes in the regulatory constraint at any given expected penalty structure, or alternatively, in the structure of the expected penalty function at any given level of the regulatory constraint, may induce discontinuous responses in the regulated equilibrium output level of the firm and industry. This will occur, if at all, at particular critical combinations of penalty structure and constraint level. At all other combinations of the policy parameters a change in either parameter has no effect on the regulated equilibrium output of the firm or industry.

Returning to Figure 2-7-3, panel (ii) illustrates an expected penalty function that is locally convex but is bounded from above by some finite level $\overline{H}$. Many enabling regulatory statutes contain maximum penalties that may be imposed. Given that the probability of detection is bounded, then so too, in these cases, is the expected penalty.

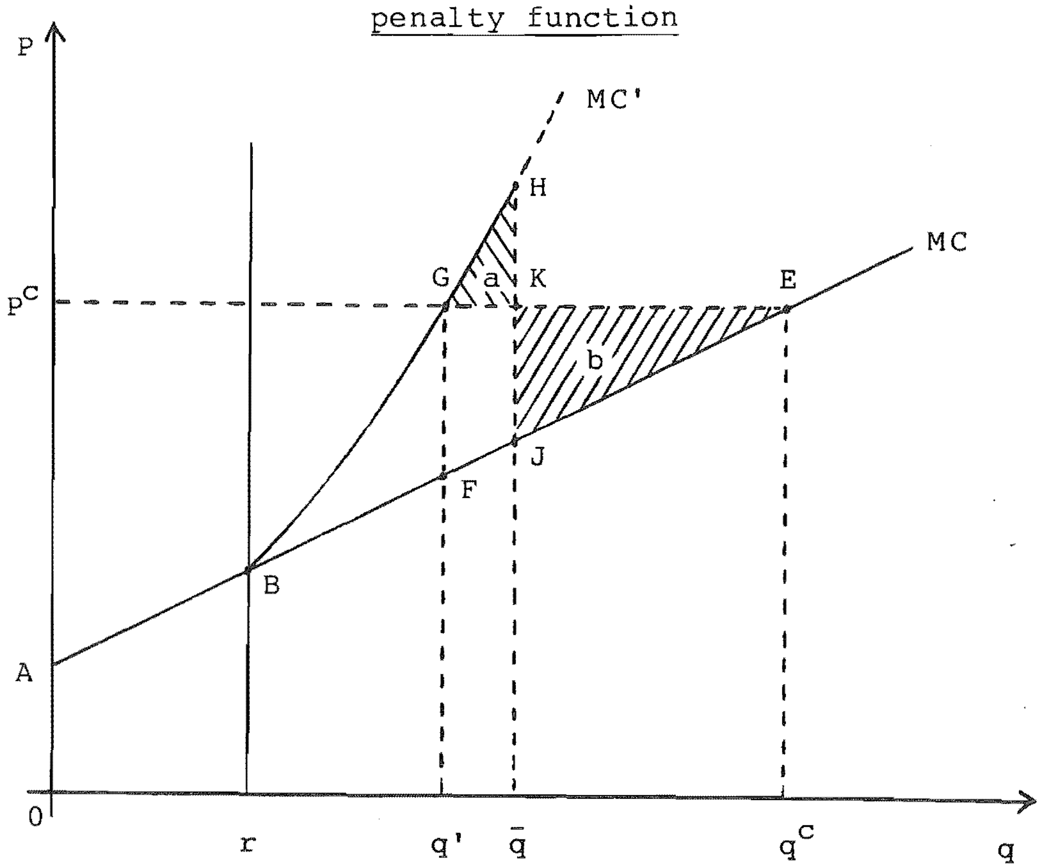
If the upper bound on the expected penalty is set at a sufficiently high level so that it is seldom, if ever, binding, then the industry operates in the output region where the expected penalty function is locally convex. In this case the expected penalty may be treated as if the firm's profit function is concave and the analysis of Section 2-6 and, in the sales tax case, Sections 2-4 and 2-5, is applicable.

This type of penalty structure however can exhibit difficulties associated with the "spillover" problem [Stigler 1970]. With a very high upper bound, the penalty may not be believable given other penalties, or marginal deterrents may be insufficient to prevent other offences of a potentially more serious nature being committed in an effort to avoid detection of the original offence.

Assuming then that the upper bound is relevant to industry decisions, a bounded convex expected penalty function may exhibit some of the properties associated with a lumpsum expected penalty. This can be shown with reference to Figure 2-7-6 below.

The firm's supply curve in the unregulated environment is given by MC. In the context of Section 2-6, with a strictly convex expected penalty function and strictly concave profit function, the firm's regulated supply curve is MC' running through points ABGH. This would give the firm's optimal regulated output level, consistent with price $P^c$ and quota r, at q' where (2-6-5) is satisfied with equality.

Figure 2-7-7: The effect of an output quota enforced by a bounded locally convex expected penalty function




In the present context, the firm's profit function is not globally concave and hence q' associated with point G need not represent a global profit-maximizing output level at price $P^c$ and quota level r. To see this area 'a' and 'b' must be compared. It is assumed in the diagram that, given the quota level r, the expected penalty reaches its maximum value, shown by area BHJ, at output level $\bar{q}$.

Beginning from q' at G, an increase in output to E has an ambiguous effect on overall profitability. The initial $q'\bar{q}$ units reduce profit from its level at q' by area GHK denoted by 'a'. The subsequent $\bar{q}q^c$ units however increase profit by KEJ denoted by area 'b'. The global profit-maximizing output level for the firm at price $P^c$ occurs at

$q^c$ when $b > a$ and $q'$ when $b < a$.

Assume initially that the profit-maximizing output level is at G in accordance with (2-6-5). If the quota level is reduced from r without adjustment to the upper bound on the expected penalty, area 'b' will increase while area 'a' will decrease. The smaller is the quota associated with any given bounded expected penalty function the more likely it is that the quota will be disregarded. In particular it is possible that the act of attempting to tighten control on the industry by reducing the available quota will suddenly induce a leap back to the unregulated equilibrium output level.

Conversely an increase in the upper bound on the expected penalty at any quota level acts to increase the size of area 'a' in Figure 2-7-6 and reduce 'b'. Increasing the upper bound from a situation where the quota has no effect on market equilibrium can therefore suddenly induce a leap in industry output from the unregulated competitive equilibrium level to that associated with an unbounded convex expected penalty given the quota r.

Following Proposition 2-7-1 similar results hold when the industry is regulated by a sales tax with a bounded convex expected penalty function. Unless the penalty function incorporates an initial unit rate component that is independent of the violation, full declaration in the case of a sales tax or full compliance with an output quota cannot be achieved. The results of the above analysis are now summarized.

PROPOSITION 2-7-6: Suppose that an industry comprising a finite number of identical competitive firms, which have concave profit functions in the unregulated environment, faces a market demand curve with non-zero finite elasticity and is subject to either a strictly positive unit rate sales tax or strictly binding tradeable output quota enforced by means of a bounded expected penalty function that is otherwise convex in output. Optimal firm and industry behaviour in the regulated environment is determined by global profit maximization which may or may not produce outcomes consistent with first-order conditions. In particular it is possible that changes in the regulatory instrument, at any given bounded expected penalty function for which the upper bound is relevant to industry behaviour, will induce discontinuous responses in regulated equilibrium output levels. Thus sufficiently high tax rates or sufficiently small output quotas may result in the regulated industry equilibrium coinciding with the unregulated competitive equilibrium level. Similarly, changes in the upper bound of the expected penalty at any value of the regulatory constraint may also induce discontinuous

<u>responses in firm and industry behaviour and hence in regulated equilibrium output levels.</u>

Given that many penalty structures are a composite of fixed and variable components and enabling legislations often stipulate maximum fines, it appears from the analysis contained in this section that discontinuity and other perverse responses to parameter changes are likely to occur in a regulated environment. This suggests that the ability of a regulatory body to control an industry in any systematic manner is severely limited. However, the literature provides evidence consistent with the intuitively appealing deterrence results presented in Sections 2-4 to 2-6 [Ehrlich, 1973; Epple and Visscher, 1984] and hence the implications of enforcement within the context of a non-concave expected penalty function is further explored in the following sections.

Finally it must be noted that for all forms of the expected penalty function considered here, deterrence requires the use of scarce, otherwise productive resources in enforcement. Recalling (2-4-10), given that both the probability term and the monetary fine, over the relevant output range, are finite, the expected penalty if either component is zero is also zero. If no resources are devoted to enforcement activities, the probability of detection is zero and no effective deterrent exists.

This was not always recognized in the early literature on the optimal structure of expected penalty functions. Following Becker's seminal paper on the economics of crime [Becker, 1968], the analysis of the optimal penalty was extended to produce an optimal fine that is infinitely large

[Carr - Hill and Stern, 1979; 303]. From limit theory, given an infinite fine, any positive expected penalty can be generated with a zero probability of detection, the precise value depending on the rates of change of its component parts. This leads to an "optimal penalty" structure which comprises an infinite monetary penalty and zero probability of detection.

This result, which is motivated by a desire to minimize the social cost of offences and is assisted by the assumption of risk aversion, is theoretically appealing but has little empirical validity as has been demonstrated in the more recent literature where the concept of a lower bound "threshold probability" sufficient to ensure deterrence is introduced [Block and Lind, 1975; Polinsky and Shavell, 1979; Pyle, 1983].

Often, as was mentioned above, enabling legislation for regulation provides for a maximum penalty that may be imposed. This maximum fine is usually set so as to maintain marginal deterrents sufficient to prevent spillover effects between different classes of offence. Even if this is not so, the constraints of market conditions will inevitably result in the industry and hence the individual firm producing finite output levels in equilibrium. For any finite output, even given an expected penalty function that is strictly convex in output and violation size, the expected penalty is also finite.

In these circumstances the "optimal penalty" result could be supported if the assumption was made that informational differences create misperceptions on the part of individual agents leading to pessimistically incorrect estimates of the probability term. This assumption seems unreasonable in the context of rational maximizing agents, at least in the longer run, with any rational- or adaptive-expectation forming process. It must be assumed therefore that, with a finite monetary fine, any effective deterrent involves the employment of scarce otherwise productive resources. This represents a cost to the economy of any regulation which has important implications in the consideration of regulatory objectives in the following Chapter.

## 2-8 CONCLUSION

This chapter has dealt with the regulation of a competitive, negative-externality-generating industry in a partial equilibrium framework. The concept of a regulated equilibrium was defined and the effects on this equilibrium of changes in policy variables were derived. Specific consideration and treatment of the necessity for enforcement confirmed the major results of the deterrence literature and showed that they depend crucially on the form of the expected penalty function. In particular, the examination of specific forms of expected penalty function in Section 2-7 revealed that the deterrence capabilities of the regulator are severely limited in certain cases.

The assumption of global concavity of firms' profit functions in the regulated environment, which is used in the majority of the chapter, facilitates continuity of behaviour. As such it provides a useful framework for analysing the effects of regulation on the industry and the responses of regulatory policy to parameter changes. This assumption however is not necessarily valid in enforcement practice.

In many instances, enabling legislation contains maximum penalties which may be imposed on offenders. The existence of a maximum fine implies that the expected penalty is bounded from above which invalidates the assumption of globally concave profit functions. With this form of penalty, as is shown in Section 2-7, local optima are not necessarily globally optimal and hence industry behaviour within the regulated environment is likely to exhibit discontinuities.

Given enforcement by means of an expected monetary fine, the analysis showed the policies employing traditional "price-oriented" instruments, such as a sales tax, could be designed to have identical aggregate outcomes to those using "quantitative" instruments such as output quota. In each case the marginal expected penalty is the per unit "price" of engaging in illegal activity. It is this price which supports the existence of either regulatory instrument. When illegal behaviour and enforcement are allowed for then, there is a sense in which any form of regulatory control works fundamentally through the pricing mechanism.

There are differences between the output quota and sales tax however. These stem from the fact that the existence of a non-zero output quota reduces the number of units, at any given output level which exceeds the quota, on which the "price" for illegal behaviour must be paid.

To the extent that a change in the amount of available quota alters this price, the regulated equilibrium output level of the industry is affected. With an expected penalty function that is strictly convex in the extent of illegal output and a fixed level of enforcement, a reduction in available quota increases the marginal expected penalty with respect to output at any given output level which exceeds the quota and hence, following Proposition 2-6-2, leads to a decrease in the regulated equilibrium output level of the industry. Given that the marginal expected penalty with respect to output varies directly with the level of enforcement, any regulated equilibrium output level can be achieved by a smaller quota coupled with a lower level of enforcement. This suggests that the costs of regulation are not independent of the type and level of the regulatory instrument used.

The other major difference between an output quota and a sales tax is in their distributional effects. Under a sales tax, agents can avoid the price of illegal behaviour only by paying the price of legal behaviour; that is, by incurring tax liability. With a tradeable output quota, however, holders of the quota are in effect paid the penalty that would have otherwise been incurred if those units had to be produced illegally. These payments accrue to agents initially allocated quotas in the form of either implicit rentals to those who are members of the industry or explicit windfall gains to those who are not.

The distributional and cost-effectiveness aspects of regulatory policy become important when questions concerning optimal regulation are raised. These issues are discussed in the following Chapter in the context of two competing and conflicting theories of economic regulation, the Naive Public Interest Theory and the Capture Theory.

NOTES

1. In the discussion that follows, the government and the regulator are assumed to be synonymous bodies.
2. See for example Baumal and Oates,1981; Buchanan and Tullock,1975; Harford,1978; Migué,1977; Montgomery,1972; Papps,1985; Storey and McCabe,1980; and Weitzman,1974 and 1978.
3. The pitfalls of using this measure are well known [Hausman, 1981; 663]. For consumer surplus to represent an exact measure of welfare change requires, among other things, that consumers' utility functions are quasi-linear and, in a two good world, that both commodities are consumed [Manning, 1986; 92]. Given the assumptions sufficient to ensure consistent aggregation in (2-2-2), this condition will not be satisfied [Manning, 1986; 98] but nevertheless, in the context of the present analysis, the use of consumer surplus represents a tolerable simplifying assumption [Willig, 1977].
4. In most diagrams demand and supply schedules are drawn as linear functions for graphical simplicity. This does not affect the qualitative nature of the illustrated results.
5. Penalizing consumers of illegal commodities is not uncommon. This would in general depress the demand side of the market and hence reduce the scope for regulating producers. It is assumed here, however, that producers represent a more compact group than consumers and thus pose fewer difficulties from the perspective of enforcement.

6. See for example Becker, 1968; Block and Lind, 1975; Carr-Hill and Stern, 1979; Cloninger, 1975; Erhlich, 1972 and 1973; Epple and Visscher, 1984; Harford, 1978; Mathur, 1978; and Polinsky and Shavell, 1979.

7. This assumes that the quota is initially allocated only to firms within this industry.

8. The precise nature of this shift is as yet unspecified and is drawn as shown in Figure 2-6-2 for illustrative purposes only.

# CHAPTER THREE

## THE OBJECTIVES OF THE REGULATOR: IMPLICATIONS FOR POLICY IN PARTIAL EQUILIBRIUM

### 3-1 INTRODUCTION

The analysis of regulated equilibrium in Sections 2-4 to 2-7 presented the effects of regulatory control on firm, and industry, behaviour and illustrated several possible outcomes. It was shown that the aggregate outcome and, in particular, the distributional effects of a regulation, depend on the instrument used and the form of the expected penalty function.

From an examination of the implications of Sections 2-4, 2-5 and 2-6, it is clear that the regulator can design a tax/expected-penalty or quota/expected-penalty policy mix that will result in identical output levels, both for the industry and individual firms, which may or may not be at the first-best output $\hat{Q}$ in Figure 2-2-2.

If the aggregate outcome is independent of the instrument used but not of the instrument/expected-penalty mix, and any output level not exceeding the unregulated competitive equilibrium level is technically feasible with a non-negative expected penalty, the actual policy employed must be dependent on something other than a particular output objective.

In Chapter One several theories of regulation were outlined. Of these two are examined here; the Naive Public

Interest Theory (NPIT) and the Capture Theory (CT).

## 3-2 NAIVE PUBLIC INTEREST THEORY: OPTIMAL POLICY

Under NPIT the regulator is motivated by a desire to achieve the maximization of social welfare. In the context of the present partial equilibrium analysis the regulator acts to maximize aggregate social surplus generated in this industry. In Figure 2-2-2, assuming costless implementation of the regulation, this occurs at $\hat{Q}$ where social surplus is shown by area $A\hat{E}H$.

At this stage it must be recalled that implementation of a regulation is not costless because of the necessity for enforcement which, as stated in Section 2-7, requires the employment of scarce, otherwise productive, resources.

The expected penalty function in both the sales tax and output quota cases has hitherto been portrayed as a non-decreasing function of the firm's output and degree of constraint violation, given values of the parameters $\alpha$ and $\beta$ respectively. Here these parameters are taken to represent the size of resources devoted to enforcement activities denoted by $L_e$.

(3-2-1) $$0 \leq L_e \leq \bar{L}_e$$

where $\bar{L}_e$ is the minimum level of enforcement activity that ensures a probability of detection of unity. If $L_e = 0$ and no resources are devoted to enforcement, the probability of detection is zero and hence, from (2-4-12) and (2-6-3), the expected penalty at all levels of illegal activity is also zero.

The cost of enforcement activity ($\omega$) is the opportunity cost of the resources used in the enforcement process which is assumed to be strictly positive with an increasing marginal cost, thus

(3-2-2) $\qquad \omega = \omega(L_e),\ \omega(0) = 0,\ \omega' > 0,\ \omega'' > 0$

The benefit from enforcement is the increase in the value of the regulator's objective function which results from the increase in industry compliance with the regulation that enforcement engenders. For a NPIT regulator, the benefit from enforcement is the increase in aggregate surplus which it generates.

Beginning, in Figure 2-2-2, from the competitive equilibrium output level at $Q^C$, the maximum increase in aggregate social surplus is shown by area $E^C B\hat{E}$ and occurs when output is restricted to the first-best level $\hat{Q}$. The marginal benefit of a unit reduction in output caused by enforcement is given by the difference between the demand price, which shows the marginal social benefit of output, and the marginal social cost, at any given output level. In Figure 2-2-2 this value is maximized at $E^C B$ which occurs at the competitive equilibrium output level $Q^C$ and, given the assumptions which underly the diagram, is strictly decreasing to zero as output falls to $\hat{Q}$.

With a non-decreasing marginal cost and strictly decreasing marginal benefit of enforcement there is a unique optimal amount of enforcement effort. It may be that, because of prohibitive costs, the optimal amount of enforcement is zero and hence the unregulated competitive equilibrium output level is the socially optimal regulated output even though it involves a degree of lost potential surplus. To preclude this possibility it is assumed that

(3-2-3) $\quad \omega'(0) < MSC(Q^C) - MSB(Q^C)$

Given the conditions on the cost of enforcement assumed in (3-2-2), this ensures that the NPIT optimal regulated output level Q* occurs at some intermediate point,

(3-2-4) $\quad \hat{Q} < Q^* < Q^C$

The determination of Q* is derived in the following analysis.

From the analysis of Sections 2-4, 2-5 and 2-6, it follows that equilibrium industry output in the regulated environment is a function of the regulatory instrument used and, through the expected penalty function, of the level of enforcement. Thus

(3-2-5) $\quad Q = Q(L_e, \Omega, Z) \quad ; \quad \frac{\partial Q}{\partial L_e} \leq 0, \ \frac{\partial Q}{\partial \Omega} \gtrless 0$

where $\Omega$ is a proxy for the size of the regulatory instrument. It some cases therefore $\Omega$ is a tax rate while in others it represents a quota level. The properties of the function Q(.) depend on the instrument used and upon the expected penalty function. The regulated equilibrium is defined for fixed values of some as yet unspecified parameters contained in the parameter vector Z. Various factors affect the supply and/or demand for the regulated commodity and are therefore relevant to the decisions of the regulator. Factors to be considered here include the income level of consumers, technological improvements which reduce the extent of externality generated by the industry, and the state of enforcement technology.

In the case of an output quota, following Proposition 2-6-2, industry equilibrium output is assumed to be monotonically decreasing in $L_e$ provided that output exceeds the quota. Once full compliance is attained, any further increase has no effect on behaviour and thus $\partial Q/\partial L_e = 0$. From Proposition 2-5-1, industry equilibrium output is assumed to be monotonically decreasing in $L_e$ in the case of a sales tax provided that the marginal expected penalty with respect to output, given that no output is declared, at the regulated equilibrium output level is less than the tax rate. If however $G_Q(Q,0,\alpha) \geq t$, $G_q = - G_x$ and $G_{qq} = G_{xx} = - G_{xq}$, then $\partial Q/\partial L_e = 0$.

Propositions 2-5-1 and 2-6-2 also show how a change in the value of the regulatory instrument affects regulated equilibrium output. In the case of a sales tax it is assumed, following Proposition 2-5-1, that an increase in the tax rate reduces regulated equilibrium output provided that the tax rate is less than the marginal expected penalty with respect to output, given that no output is declared, evaluated at the regulated equilibrium output level. If the tax rate exceeds this value, further increases in the tax rate do not affect behaviour and hence $\partial Q/\partial \Omega = 0$. In the case of an output quota, it is assumed, following Proposition 2-6-2, that an increase in available quota, which remains binding on behaviour, raises industry regulated equilibrium output. If the quota is enforced by a constant unit rate expected penalty, changes in the quota level which remain binding do not affect the marginal expected penalty with respect to output and hence $\partial Q/\partial \Omega = 0$.

The range of the function in (3-2-5) is restricted to the set of feasible illegal output levels given a non-zero expected penalty, thus

(3-2-6) $$0 \leq Q_{FC} \leq Q(.) \leq Q^C$$

where $Q^C$ is the unregulated competitive equilibrium level and $Q_{FC}$ the output level consistent with full compliance with the announced regulatory instrument.

From the market demand function shown in (2-2-2) the inverse demand function, showing the demand price at any regulated equilibrium quantity given values of the relevant parameters, can be derived. Thus

(3-2-7) $$P = h(Q(L_e,\Omega,\mu,Y),Y) \ ; \quad h_1 < 0, \ h_2 > 0$$

where Y represents aggregate consumer income and $\mu$ the state of enforcement technology. It is assumed that the demand curve is negatively sloped and that the regulated commodity is a normal good. The enforcement-exclusive social cost of production is given by

(3-2-8) $$SC = SC(Q(L_e,\Omega,\mu,Y),\phi) \ ; \ SC_1 > 0, \ SC_{11} > 0, \ SC_2 < 0$$

where $\mu$ and Y are as defined in (3-2-7) and $\phi$ represents the state of technology in the regulated industry which determines the degree of externality that the industry generates. Following the analysis of Chapter Two, it is assumed that enforcement-exclusive social cost is strictly convex in output. It is also assumed that an increase in $\phi$ represents an improvement in technology in that it reduces the extent of externality generated in the industry. The derivation and properties of (3-2-7) and (3-2-8) are further explained in Appendix 3-1.

The NPIT regulator, in maximizing the social surplus generated in the industry, seeks to solve the following expression.

$$(3\text{-}2\text{-}9)\quad \underset{L_e,\Omega}{\text{Maximize}}\; W_{R_e} = \int_{Q^C}^{Q(L_e,\Omega,\mu,Y)} [h(Q,Y) - SC_1(Q,\phi)]dQ - \omega(L_e)$$

In terms of Figure 2-2-2, regulating the industry is equivalent to reducing industry output from $Q^C$. The payoff to regulatory activity is then the recovery of the lost potential aggregate surplus that exists at $Q^C$ when compared with the first-best outcome at $\hat{Q}$. This is shown by the integral term of (3-2-9). The second term is the resource cost of enforcing the regulation.

The first-order conditions of (3-2-9) are[1]

$$(3\text{-}2\text{-}10)\quad \frac{\partial W_{R_e}}{\partial L_e} = [h(Q(L_e^*,\Omega^*,Z),Y) - SC_1(Q(L_e^*,\Omega^*,Z),\phi)]\frac{\partial Q}{\partial L_e} - \omega'(L_e^*) \leq 0 \quad ; \quad < \text{ only if } L_e^* = 0$$

$$(3\text{-}2\text{-}11)\quad \frac{\partial W_{R_e}}{\partial \Omega} = [h(Q(L_e^*,\Omega^*,Z),Y) - SC_1(Q(L_e^*,\Omega^*,Z),\phi)]\frac{\partial Q}{\partial \Omega} \leq 0 \quad ; \quad < \text{ only if } \Omega^* = 0$$

The first term of (3-2-10) is the difference between the marginal social cost of production and the demand price, or marginal social benefit of output, at the optimal level of enforcement, multiplied by the marginal effect of enforcement on output. This gives the marginal benefit of enforcement while the second term is the marginal cost of an additional unit of enforcement. At the optimum level of enforcement, assuming that enforcement takes place, these are equated. If the marginal cost of enforcement everywhere exceeds its marginal benefit, no enforcement will occur.

Conditions (3-2-10) and (3-2-11) must be satisfied simultaneously at the optimal NPIT regulated equilibrium output level $Q^* = Q(L_e^*, \Omega^*, Z)$.

LEMMA 3-2-1: Suppose an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax enforced by means of an expected monetary penalty. Assuming that an increase in enforcement reduces the equilibrium output level of the industry and that the marginal cost of enforcement is strictly positive, at the NPIT-optimal regulated industry equilibrium output level the marginal social cost of output strictly exceeds the demand price.

Proof.

This follows immediately from (3-2-10). Given that $\omega'(L_e) > 0$ and that $\partial Q/\partial L_e < 0$, then, if any enforcement is to occur under the NPIT framework so that $L_e^* > 0$, the term $[h(Q(L_e^*,\Omega^*,Z),Y) - SC_1(Q(L_e^*,\Omega^*,Z),\phi)]$ must be negative. If, however, $L_e^* = 0$ then no deterrent to non-compliance with the regulation exists and hence the regulated equilibrium coincides with the unregulated competitive equilibrium.

Recalling the assumption that the unregulated competitive equilibrium output level in this externality-generating industry is socially excessive, the term $[h(Q(L_e^*,\Omega^*,Z),Y) - SC_1(Q(L_e^*,\Omega^*,Z),\phi)]$ is negative in this case also. Hence, at any feasible NPIT-optimal regulated equilibrium output level, the marginal social cost of output strictly exceeds the demand price.

□

Using Lemma 3-2-1 in (3-2-11), the optimal value of the regulatory instrument $\Omega$ occurs when $\frac{\partial Q}{\partial \Omega} = 0$. If, however, an increase in the value of the regulatory instruments always acts to increase regulated equilibrium output, the optimal value of the instrument is zero.

<u>PROPOSITION 3-2-1</u>: <u>Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly positive unit rate sales tax or a strictly binding output quota each of which is enforced by means of an expected monetary penalty. Given that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation and that the marginal cost of enforcement is non-decreasing, NPIT-optimal regulation of the industry exhibits the following characteristics:</u>

(i) <u>Given that the expected penalty function is binding on the industry's behaviour, the optimal amount of enforcement occurs where the marginal social benefit</u>

<u>of enforcement equals its marginal social cost and is uniquely determined giving a regulated equilibrium output level which exceeds the first-best optimum</u>.

(ii) <u>The optimal quota is zero. This is not always uniquely optimal however. For instance, if the quota is enforced by a constant unit rate expected penalty, the level of the quota is irrelevant to the NPIT regulator provided that it remains binding</u>.

(iii) <u>In the case of regulation by a unit rate sales tax the optimal tax rate will be set at a rate no less than that which equates the difference between market price and the marginal private cost of production at the output level consistent with part (i) of the Proposition</u>.

<u>Proof</u>.

(i) The assumptions that the expected penalty is binding on behaviour and that the firm's profit function is strictly concave ensures, following Proposition 2-5-1, that $\partial Q/\partial L_e < 0$. The result then follows using (3-2-2), (3-2-3) and Lemma 3-2-1.

(ii) Given that, from part (i) of the Proposition, it is optimal to control the industry so that $L_e^* > 0$, there is some level of quota strictly less than the unregulated competitive output level that is optimal for the NPIT regulator. This is obvious from the realization that enforcement can be optimal only if there is some binding constraint to enforce.

Following Proposition 2-6-2, with an expected penalty function that is convex in output and the extent of constraint violation, $\partial Q/\partial\Omega > 0$ and any reduction in available quota reduces regulated equilibrium output. Thus, using Lemma 3-2-1 in (3-2-11), at any strictly positive quota level $\partial W_{R_e}/\partial\Omega < 0$ and hence the optimal quota level is zero.

If the expected penalty is a constant unit rate fine, then a change in quota level, providing that the quota remains binding on behaviour, does not affect the marginal expected penalty with respect to output and hence, from Proposition 2-6-2, $\partial Q/\partial\Omega = 0$. Thus in (3-2-11), $\partial W_{R_e}/\partial\Omega = 0$ at any binding quota level and hence, providing that the quota is binding, the level of quota is irrelevant to the NPIT regulator.

(iii) By the same reasoning as in part (ii) of the Proposition, given that, from part (i) of the Proposition, it is optimal to control the industry so that $L_e^* > 0$, there is some strictly positive tax rate that is optimal for the NPIT regulator.

Following Proposition 2-5-1, with an expected penalty function that is convex in output and the extent of constraint violation, $\partial Q/\partial\Omega < 0$ provided that the tax rate is less than the marginal expected penalty with respect to output, given that no output is declared, evaluated at the equilibrium output level that is consistent with full compliance with the tax.

Thus for any tax rate $\bar{t}$ such that

$$\bar{t} = h(Q(L_e^*,\bar{t},Z),Y) - MPC(Q(L_e^*,\bar{t},Z)) \tag{3-2-12}$$
$$< G_Q(Q(L_e^*,\bar{t},Z),0,L_e^*)$$

using Lemma 3-2-1 in (3-2-11) gives $\partial W_{R_e}/\partial\Omega > 0$ and it is optimal for the NPIT regulator to increase the tax rate. This is the case for all tax rates such that $\partial Q/\partial\Omega < 0$.

From Proposition 2-5-1, $\partial Q/\partial\Omega = 0$ for all tax rates $t \geq \bar{\bar{t}}$ where

$$\bar{\bar{t}} = h(Q(L_e^*,\bar{\bar{t}},Z),Y) - MPC(Q(L_e^*,\bar{\bar{t}},Z)) \tag{3-2-13}$$
$$= G_Q(L_e^*,\bar{\bar{t}},Z),0,L_e^*)$$

From (3-2-11) therefore, using Lemma 3-2-1, all tax rates $t \geq \bar{\bar{t}}$ are optimal under NPIT regulation. Tax rate $\bar{\bar{t}}$, which equals the difference between the demand price and marginal private cost of production evaluated at the equilibrium output level consistent with the marginal expected penalty, given that no output is declared, generated by the optimal enforcement level $L_e^*$, represents the lower bound on the NPIT optimal unit rate of sales tax.

□

Proposition 3-2-1 is intuitively appealing. The first result, that optimal enforcement from a public interest objective occurs where its marginal social benefit is equal to its marginal social benefit, is widely found, in similar forms, in the deterrence literature.[2] The most significant implication of this result is that, with an enforcement cost which is positive and increasing in enforcement activity, a NPIT regulator will never seek to restrict the output of an externality generating industry to the first-best level.

Results (ii) and (iii) of Proposition 3-2-1 show that, in the case of a sales tax or output quota with an expected penalty which is convex in output and the extent of constraint violation, the optimal level of enforcement, and hence the optimal degree of control exercised on the industry, is independent of the regulatory instrument used. The optimal level of the regulatory instrument itself however is not.

Result (ii) suggests that a NPIT regulator will never choose to regulate the industry by strictly positive output quota given the assumption of convexity of the expected penalty function. This is consistent with economists' general preference for "price-oriented" restrictions such as taxes rather than quantitative controls.

Given that no tax is voluntarily complied with[3] result (iii) shows that any tax that is fully complied with in the enforced regulatory environment is too low to be optimal for a NPIT regulator. In fact the NPIT-optimal tax rate is set so that, given the optimal amount of enforcement and hence the optimal regulated equilibrium output level, at most only the marginal unit of output is declared.

In each case it is the expected penalty that is directly binding on behaviour. The expected penalty then represents an unavoidable tax on the actions of the offending firms and industry and it is through the marginal effects of this "tax" that behaviour is modified. The problem of the NPIT regulator is therefore to select the appropriate level of the "tax" given the resource costs of its implementation. In both part (ii) and part (iii) of Proposition 3-2-1, the value of the regulatory instrument is set so as to maximize the extent of constraint violation at any given output level. Following the assumption of the convexity of the expected penalty in violation size and output, this action minimizes the resource cost of enforcement necessary to generate an expected penalty that is sufficient to restrict output to any given level.

## 3-3 NAIVE PUBLIC INTEREST THEORY: POLICY RESPONSES TO PARAMETER CHANGES

The problem of the NPIT regulator given in (3-2-9) is implicitly defined for some fixed value of the parameter vector Z and of the other parameters which do not appear in Z. Thus the first-order conditions (3-2-10) and (3-2-11) provide the optimal amount of enforcement and level of the regulatory instrument as functions of these parameters.

Assuming that $W_{R_e}$ is concave, solving (3-2-10) and (3-2-11) simultaneously gives the optimal values of enforcement and the level of the regulatory instrument

$$(3\text{-}3\text{-}1) \qquad L_e^* = L_e^*(\mu,\phi,Y) \; ; \; \Omega^* = \Omega^*(\mu,\phi,Y)$$

where $\mu$ represents the state of enforcement technology, $\phi$ the state of technology in the regulated industry and Y the aggregate level of consumer income.

Totally differentiating (3-2-10) and (3-2-11) allows the use of comparative static analysis to examine the qualitative effects of changes in these parameters on the NPIT regulator's choice of optimal regulatory policy.

From (3-2-10)

$$(3\text{-}3\text{-}2) \qquad AdL_e + Bd\Omega + Ed\mu + FdY - Jd\phi = 0$$

where

$$A = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e{}^2} + [h_1(.)\frac{\partial Q}{\partial L_e} - SC_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial L_e} - \omega''(L_e)$$

$$B = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \Omega} + [h_1(.)\frac{\partial Q}{\partial \Omega} - SC_{11}(.)\frac{\partial Q}{\partial \Omega}]\frac{\partial Q}{\partial L_e}$$

$$E = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \mu} + [h_1(.)\frac{\partial Q}{\partial \mu} - SC_{11}(.)\frac{\partial Q}{\partial \mu}]\frac{\partial Q}{\partial L_e}$$

$$F = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial Y} + [h_1(.)\frac{\partial Q}{\partial Y} + h_2(.) - SC_{11}(.)\frac{\partial Q}{\partial Y}]\frac{\partial Q}{\partial L_e}$$

$$J = SC_{12}(.)\frac{\partial Q}{\partial L_e}$$

and from (3-2-11)

$$(3\text{-}3\text{-}3) \qquad MdL_e + Nd\Omega + Sd\mu + TdY - Ud\phi = 0$$

where

$$M = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial \Omega \partial L_e} + [h_1(.)\frac{\partial Q}{\partial L_e} - SC_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial \Omega}$$

$$N = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial \Omega^2} + [h_1(.)\frac{\partial Q}{\partial \Omega} - SC_{11}(.)\frac{\partial Q}{\partial \Omega}]\frac{\partial Q}{\partial \Omega}$$

$$S = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial \Omega \partial \mu} + [h_1(.)\frac{\partial Q}{\partial \mu} - SC_{11}(.)\frac{\partial Q}{\partial \mu}]\frac{\partial Q}{\partial \Omega}$$

$$T = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial\Omega\partial Y} + [h_1(.)\frac{\partial Q}{\partial Y} + h_2(.)-SC_{11}(.)\frac{\partial Q}{\partial Y}]\frac{\partial Q}{\partial\Omega}$$

$$U = SC_{12}(.)\frac{\partial Q}{\partial\Omega}$$

Rearranging (3-3-2) and (3-3-3) in matrix form and solving

$$(3\text{-}3\text{-}4) \qquad \begin{bmatrix} dL_e \\ d\Omega \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} N & -B \\ -M & A \end{bmatrix}\begin{bmatrix} -Ed\mu - FdY + Jd\phi \\ -Sd\mu - TdY + Ud\phi \end{bmatrix}$$

where $\begin{bmatrix} A & B \\ M & N \end{bmatrix}$ is the 2 x 2 matrix of coefficients of the terms in $dL_e$ and $d\Omega$ from (3-3-2) and (3-3-3) and

$$(3\text{-}3\text{-}5) \qquad \Delta = AN - BM > 0$$

is the determinant of this coefficient matrix which is positive by the assumption of concavity of (3-2-9). In addition, concavity implies that A and hence N are negative.

The first parameter change considered concerns the level of aggregate consumer income.

(1) <u>The Level of Aggregate Consumer Income</u>

<u>PROPOSITION 3-3-1:</u> <u>Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly positive unit rate sales tax or a strictly binding output quota each of which is enforced by means of an expected monetary penalty that is strictly convex in output and the extent of constraint violation. Suppose also that regulating the output</u>

of the industry necessitates the use of scarce resources to enforce the regulation, that the marginal cost of enforcement is non-decreasing so that the NPIT regulator's objective function is strictly concave, and hence that the NPIT-optimal regulatory policy at some fixed vector of parameter values is in accordance with the results of Proposition 3-2-1:

(i) Suppose that a change in aggregate consumer income leaves the marginal productivity of the regulator's policy instruments unchanged. An increase in aggregate consumer income, which increases demand for the output of the regulated commodity, reduces (increases) the NPIT-optimal level of enforcement and reduces (increases) the NPIT-optimal tax rate if the income-induced increase in demand price at the initial regulated equilibrium exceeds (is less than) the income-induced change in the per unit loss of aggregate surplus on the marginal unit of output.

(ii) Suppose that an increase in aggregate consumer income, which

<u>increases demand for the output of the regulated industry, improves (reduces) the marginal productivity of the regulator's policy instruments. If the income-induced rise in demand price at the initial regulated equilibrium is less than (exceeds) the income-induced change in the per unit loss of aggregate surplus on the marginal unit of output, the increase in income raises (lowers) the NPIT-optimal tax rate and increases (reduces) the NPIT-optimal level of enforcement. If, however, the income-induced increase in demand price exceeds (is less than) the change in the unit rate loss of aggregate surplus, the effect of the increase in income on NPIT-optimal regulatory policy is ambiguous.</u>

<u>Proof</u>.

(i) Using (3-3-4) with $d\mu = d\phi = 0$ gives

(3-3-5) $$\frac{dL_e^*}{dY} = -\frac{N}{\Delta}\cdot F + \frac{B}{\Delta}\cdot T$$

In the case of a sales tax, following Proposition 2-5-1, $\partial Q/\partial L_e < 0$ and $\partial Q/\partial \Omega \leq 0$, while, from the assumption of concavity of (3-2-9), $\Delta > 0$ and $N < 0$. The discussion of the component terms of (3-3-4) in in Appendix 3-2 shows in addition that $B > 0$. Using this information in (3-3-5) reveals that

$$\frac{dL_e^*}{dY} \gtrless 0 \quad \text{if } F,T \gtrless 0 \tag{3-3-6}$$

Given the assumption that a change in aggregate consumer income does not affect the marginal productivity of the regulator's policy instruments, the component terms $\partial^2 Q/\partial L_e \partial Y$ and $\partial^2 Q/\partial \Omega \partial Y$, from F and T as defined in (3-3-2) and (3-3-3) respectively, are both zero. Following Appendix 3-1, the increase in income acts to increase the regulated equilibrium output of the industry and thus $\partial Q/\partial Y > 0$. From (3-3-2) and (3-3-3) then, assuming $\partial Q/\partial \Omega < 0$,[4]

$$F,T \gtrless 0 \quad \text{if} \quad h_2(.) \gtrless -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y} \tag{3-3-7}$$

The term $h_2(.)$ denotes, following (3-2-7), the income-induced change in demand price at the initial regulated equilibrium. The term $[h_1(.)-SC_{11}(.)]$ shows the unit rate of change in the difference between demand price $h(.)$ and marginal social cost $SC_1(.)$. Given that, from Lemma 3-2-1, the demand price is less than marginal social cost over the relevant output range, multiplying this square-bracketed term by $\partial Q/\partial Y$ gives the income-induced change in the unit loss of social surplus on the marginal unit of output. Applying the conditions from (3-3-7) in (3-3-6), it is clear that the result of part (i) of the Proposition, concerning income-induced changes in the NPIT-optimal level of enforcement, holds in the case of a unit rate sales tax.

In the case of an output quota, following Proposition 2-6-2, $\partial Q/\partial L_e < 0$ and $\partial Q/\partial \Omega > 0$ while the conditions on N and $\Delta$ remain from the concavity of the objective function. From Appendix 3-1, as a reduction in quota is qualitatively equivalent to an increase in the tax rate, $B < 0$. Using this information in (3-3-5) shows that

$$\text{(3-3-8)} \qquad \frac{dL_e^*}{dY} \gtrless 0 \text{ if } F \gtrless 0 \text{ and } T \lessgtr 0$$

Given the assumption that $\partial^2 Q/\partial L_e \partial Y = \partial^2 Q/\partial \Omega \partial Y = 0$, and that $\partial Q/\partial L_e$ and $\partial Q/\partial \Omega$ are of opposite signs as indicated above, then from (3-3-2) and (3-3-3)

$$\text{(3-3-9)} \qquad F \gtrless 0 \text{ and } T \lessgtr 0 \quad \text{if} \quad h_2(.) \lessgtr -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$$

The right-hand side of condition (3-3-9) is the same as that of (3-3-7). Following the same arguments therefore and applying them in (3-3-8) it is clear that the result in part (i) of the Proposition concerning income-induced changes in the NPIT-optimal level of enforcement holds in the case of a binding output quota also.

Again from (3-3-4) with $d\mu = d\phi = 0$

$$\text{(3-3-10)} \qquad \frac{d\Omega^*}{dY} = \frac{M}{\Delta}.F - \frac{A}{\Delta}.T$$

In the case of a sales tax, in addition to the results outlined above, $A < 0$ by the assumption of concavity of (3-2-9) and, following the discussion of (3-3-4) in Appendix 3-2, $M > 0$. Hence in (3-3-10)

$$\text{(3-3-11)} \qquad \frac{d\Omega^*}{dY} \gtrless 0 \quad \text{if} \quad F,T \gtrless 0$$

The right-hand side of condition (3-3-11) is the same as that of (3-3-6) and hence the same conclusions apply. The result of part (i) of the Proposition concerning income-induced changes in the NPIT-optimal tax rate therefore follows.

Given that, as shown in part (ii) of Proposition 3-2-1 from the analysis of (3-2-10) and (3-2-11), the NPIT-optimal quota level is zero, it is not possible to determine the parameter-induced policy response from differentiating the relevant first-order conditions. Rather, with an expected penalty function that is convex in the extent of illegal output the enforcement/quota-size tradeoff is such that it is always optimal for a NPIT regulator to reduce the size of a strictly positive output quota. Regardless of the value of a parameter, therefore, the NPIT-optimal quota is zero and thus, in the case of regulation by output quota, $\partial\Omega^*/\partial Y = 0$.

(ii) If, in the case of regulation by sales tax, an increase in income improves the marginal productivity of the regulator's policy instruments then $\partial^2 Q/\partial L_e \partial Y$ and $\partial^2 Q/\partial\Omega\partial Y$ are both negative. Following Lemma 3-2-1, marginal social cost exceeds demand price over the relevant output range. In (3-3-2) and (3-3-3) therefore, the first composite term of both F and T is positive. Applying (3-3-7) to (3-3-6) and (3-3-10), if $h_2(.) < -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$, an increase in aggregate income leads unambiguously to an increase in the NPIT-optimal tax rate and level of enforcement. Alternatively, if an increase in income reduces the marginal productivity of the regulator's policy instruments, $\partial^2Q/\partial L_e\partial Y$ and $\partial^2Q/\partial\Omega\partial Y$ are both positive and the opposite result emerges. If, however, $h_2(.) > -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$, in each case the signs of F and T are ambiguous and hence so also is the policy impact of the income change.

In the case of an output quota an increase in income which improves the marginal productivity of the regulator's policy instruments, $\partial^2Q/\partial L_e\partial Y < 0$ and $\partial^2Q/\partial\Omega\partial Y > 0$. Following Lemma 3-2-1 therefore the first composite term of F is positive while that of T is negative. Applying (3-3-9) to (3-3-8), if $h_2(.) < -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$, an increase in aggregate income leads unambiguously to an increase in NPIT-optimal enforcement. As in the sales tax case, if the increase in income reduces the productivity of the regulatory instruments, the result is reversed. If however $h_2(.) > -[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$ the effect of the increase in income on the NPIT-optimal enforcement policy is ambiguous. Finally, as observed in the proof of part (i) of the Proposition, the NPIT-optimal quota level is unaffected by changes in parameter values.

□

The results of Proposition 3-3-1 are intuitively appealing if not immediately obvious. An increase in income which increases the demand for the output of the regulated industry may well reduce the benefit to society of restricting its output. Society is more willing to incur negative externalities associated with a commodity for which demand is high than it is when there is little aggregate demand for the product. Hence it might be expected that when demand rises the NPIT regulator seeks to reduce the extent to which the industry is restricted. As Proposition 3-3-1 shows however, this is not necessarily the case.

The income-induced increase in demand acts to increase the socially optimal output level of the industry from that shown as $\hat{Q}$ in Figure 2-2-2. This of itself suggests that the marginal social benefit of restricting the output of the industry, at any given level less than the unregulated competitive equilibrium $Q^C$, is reduced. This effect is captured by the term $h_2(.)\frac{\partial Q}{\partial L_e}$. The increase in income however also increases the regulated equilibrium output level of the industry at any given regulatory policy. Given the assumptions of a negatively sloped demand curve and strictly convex social cost, and the result of Lemma 3-2-1, the loss in aggregate surplus on the marginal unit is increased at the new equilibrium. This is shown by the term $[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial Y}$. Multiplying this term by $\partial Q/\partial L_e$ gives the change in marginal benefit to the NPIT regulator of restricting industry output whereas the term $h_2(.)\frac{\partial Q}{\partial L_e}$ gives the change in marginal cost.

Assuming that the change in income does not affect the marginal productivity of the regulatory instruments, the optimal policy response is a familiar marginal cost/ marginal benefit decision and the intuition is apparent. Changes in marginal productivity of the instruments however may reinforce the policy response or reverse it depending on the relative magnitudes of the change in marginal effectiveness of the instruments and the change in the marginal benefit of restricting output at the initial instrument productivity.

The second parameter change to be considered is that of the state of technology in the regulated industry which determines the degree of negative-externality generated. As stated in Appendix 3-1, an example of such a parameter is the state of pollution abatement technology in an industry.

(2) The State of Externality-Abatement Technology.

PROPOSITION 3-3-2: Under the assumptions in the stem of Proposition 3-3-1, an improvement in technology in the regulated industry which reduces the social cost of its activities also reduces the extent to which it is optimal for the NPIT regulator to control the industry. Thus the optimal level of enforcement is reduced and the optimal tax rate is lowered.

Proof.

Using (3-3-4) with $dY = d\mu = 0$ and substituting for J and U gives

$$(3\text{-}3\text{-}12) \quad \frac{dL_e^*}{d\phi} = \frac{N}{\Delta}[SC_{12}(.)\frac{\partial Q}{\partial L_e}] - \frac{B}{\Delta}[SC_{12}(.)\frac{\partial Q}{\partial \Omega}]$$

Following the assumption contained in (A3-1-9), $SC_{12}(.) < 0$. Hence, applying the information concerning the other component terms of (3-3-12) as stated in the proof of Proposition 3-3-1, irrespective of whether regulation is by sales tax or output quota, the effect of the introduction of new technology which lowers the social cost of output from this industry is to reduce the NPIT-optimal level of enforcement activity.

Again from (3-3-4) with $dY = d\mu = 0$ and substituting for J and U gives

$$\text{(3-3-13)} \qquad \frac{d\Omega^*}{d\phi} = -\frac{M}{\Delta}[SC_{12}(\cdot)\frac{\partial Q}{\partial L_e}] + \frac{A}{\Delta}[SC_{12}(\cdot)\frac{\partial Q}{\partial \Omega}]$$

Given the previously derived conditions on the component terms of (3-3-13), in the case of a sales tax $d\Omega^*/d\phi < 0$ and the effect of the introduction of new technology which lowers the social cost of output from the industry is to reduce the NPIT-optimal rate of sales tax.

Given that, by Proposition 2-5-1, $\partial Q/\partial L_e < 0$ and $\partial Q/\partial \Omega \leq 0$ in the case of a sales tax, and that, by Proposition 2-6-2, $\partial Q/\partial L_e < 0$ in the case of an output quota, the above policy responses correspond to a reduction in the extent of NPIT-optimal control over the industry.

$\square$

In this case, the introduction of improved externality-abating technology reduces the marginal social cost of production at every output level. In terms of Figure 2-2-2 this corresponds to an increase in the socially-optimal industry output level $\hat{Q}$. The marginal benefit to the NPIT regulator of controlling industry output to any given level which is less than the unregulated competitive equilibrium is therefore reduced. This then leads to a reduction in NPIT-optimal control of the industry through some combination of tax cuts, quota increases and enforcement cuts.

If it was allowed that the technology changes affected marginal production costs, the regulated equilibrium would then be affected and the analysis would be similar to that of the income change in Proposition 3-3-1. Assuming however that the introduction of externality-abating technology increases marginal private production costs, regulated equilibrium output at any regulatory policy combination falls. This further reduces the marginal benefit to the NPIT regulator of restricting industry output and hence reinforces the result of Proposition 3-3-2.

The final parameter to be considered in this section is the state of enforcement technology.

(3) <u>The State of Enforcement Technology</u>.

<u>PROPOSITION 3-3-3</u>: <u>Under the assumptions in the stem of Proposition 3-3-1, the following results emerge</u>:

(i) <u>The effect of an improvement in enforcement technology on regulatory policy depends on the relative sizes of its effect on the marginal productivity of the regulator's policy instruments and its effect on the regulated equilibrium at any given level of the policy instruments. If the new technology improves the marginal efficiency of enforcement and the regulatory instrument sufficiently more than it reduces</u>

the regulated equilibrium output level at any given policy combination, then it is optimal for the NPIT regulator to increase the extent to which it controls the industry. Thus the optimal level of enforcement is increased and the optimal tax rate is raised. If however the effect of the new technology on the marginal efficiency of enforcement and the regulatory instrument is sufficiently small relative to its effect on the regulated equilibrium output level, then it is optimal for the NPIT regulator to reduce enforcement and the level of the tax rate.

(ii) Suppose that the introduction of new enforcement technology raises the marginal effectiveness of an increase in the tax rate sufficiently in relation to its effect on the regulated equilibrium output level at any given regulatory policy combination so that the welfare impact of a change in the tax rate is increased. Suppose also that this increase in marginal welfare impact is

<u>sufficiently small. It is possible that the effect of the new technology is to increase the NPIT-optimal tax rate and decrease the NPIT-optimal level of resources devoted to enforcement activity.</u>

<u>Proof</u>.

(i) Using (3-3-4) with $dY = d\phi = 0$ gives

$$\frac{dL_e^*}{d\mu} = -\frac{N}{\Delta}.E + \frac{B}{\Delta}.S \tag{3-3-14}$$

Given that $N < 0$ by the concavity of (3-2-9) and that, from the discussion of (3-3-3) in Appendix 3-2, $B > 0$ in the case of a sales tax, then $dL_e^*/d\mu > 0$ if terms E and S are both positive and $dL_e^*/d\mu < 0$ if $E,S < 0$. From (A3-2-5), $E \gtrless 0$ if and only if

$$\frac{\partial^2 Q/\partial L_e\partial\mu}{\partial Q/\partial\mu} \gtrless \frac{[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}}{[h(.)-SC_1(.)]} \tag{3-3-15}$$

while from (A3-2-7), $S \gtrless 0$ if and only if

$$\frac{\partial^2 Q/\partial\Omega\partial\mu}{\partial Q/\partial\mu} \gtrless -\frac{[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial\Omega}}{[h(.)-SC_1(.)]} \tag{3-3-16}$$

From the discussion of the component terms of (3-3-15) and (3-3-16) in Appendix 3-2, the right-hand side of each expression is positive. Following (A3-1-10) it is assumed that $\frac{\partial^2 Q}{\partial\Omega\partial\mu}, \frac{\partial^2 Q}{\partial L_e\partial\mu}, \frac{\partial Q}{\partial\mu} < 0$. The terms $\partial Q/\partial L_e$ and $\partial Q/\partial\Omega$ may be interpreted as the marginal effectiveness or efficiency of each policy instrument in restricting the output of the industry. Hence $\partial^2Q/\partial L_e\partial\mu$ and $\partial^2Q/\partial\Omega\partial\mu$ represent the change in the marginal effectiveness of enforcement and the rate of sales tax respectively.

Using conditions (3-3-15) and (3-3-16) in (3-3-14) then, if the introduction of the new technology raises the marginal efficiency of the tax rate and enforcement activity sufficiently in relation to the reduction in equilibrium output which it induces at any given policy combination, so that positive inequalities hold in both (3-3-15) and (3-3-16), it is optimal for the NPIT regulator to increase the level of enforcement activity. Alternatively, if the effect of the new technology on the marginal effectiveness of the tax rate and enforcement is not sufficiently large relative to $\partial Q/\partial\mu$, the negative inequality holds in both (3-3-15) and (3-3-16) and hence it is optimal for the NPIT regulator to reduce the level of enforcement activity.

In the case of an output quota, from the discussion of (3-3-2) in Appendix 3-2, $B < 0$. Hence using (3-3-14), $dL_e^*/d\mu > 0$ if $E > 0$ and $S < 0$ and $dL_e^*/d\mu < 0$ if $E < 0$ and $S > 0$. The required condition on the term E follows from (3-3-15) and is the same as in the tax case discussed above. Thus from (A3-2-7), $S \gtrless 0$ if and only if (3-3-16) holds. Following Proposition 2-6-2 and (A3-1-10) however, both sides of (3-3-16) are negative in the case of an output quota.

From (A3-1-10), $\frac{\partial^2Q}{\partial\Omega\partial\mu} > 0$ with an output quota showing that the introduction of the new enforcement technology increases the marginal effectiveness of an output quota also. Hence, using (3-3-15) and (3-3-16) in (3-3-14), if the effect of the new technology on the marginal effectiveness of enforcement and the quota level is sufficiently large relative to $\partial Q/\partial\mu$, so that the positive inequality holds in (3-3-15) and the negative inequality holds in (3-3-16), it is optimal for the NPIT regulator to increase the level of enforcement activity. The converse result emerges with the opposite inequalities in (3-3-15) and (3-3-16).

Again from (3-3-4) with $dY = d\phi = 0$

(3-3-17) $$\frac{d\Omega^*}{d\mu} = \frac{M}{\Delta}.E - \frac{A}{\Delta}.S$$

In the case of a sales tax $M > 0$, and $A < 0$ by the concavity of (3-2-9). From (3-3-17) therefore, $d\Omega^*/d\mu > 0$ if $E,S > 0$, and $d\Omega^*/d\mu < 0$ if $E,S < 0$. These conditions are identical to those applying to (3-3-14) in the sales tax case and hence the same assumptions on the relative sizes of the increase in marginal efficiency of enforcement and tax rate compared with $\partial Q/\partial\mu$ apply here also. Thus the conditions sufficient for $dL_e^*/d\mu \gtrless 0$ also imply $d\Omega^*/d\mu \gtrless 0$ in the case of regulation by sales tax.

(ii) From (3-3-14) and (3-3-17)

(3-3-18) $$\frac{dL_e^*}{d\mu} < 0 \text{ if and only if } \frac{B}{\Delta}.S - \frac{N}{\Delta}.E < 0$$

and

(3-3-19) $$\frac{d\Omega^*}{d\mu} > 0 \text{ if and only if } \frac{M}{\Delta}.E - \frac{A}{\Delta}.S > 0$$

The assumption contained in the Proposition is that the effect of the new enforcement technology is to increase the marginal efficiency of the tax rate sufficiently so that, from (3-3-16), $S > 0$. Using this and rearranging the condition given in (3-3-18)

$$E < \frac{B}{N}.S \tag{3-3-20}$$

and manipulating the condition given in (3-3-19)

$$E > \frac{A}{M}.S \tag{3-3-21}$$

Thus $dL_e^*/d\mu < 0$ and $d\Omega^*/d\mu > 0$ if and only if both (3-3-20) and (3-3-21) are satisfied. Given that the right-hand sides of both (3-3-20) and (3-3-21) are negative, this condition is satisfied only if

$$\frac{A}{M}.S < \frac{B}{N}.S \tag{3-3-22}$$

Rearranging, (3-3-22) simplifies to $AN-BM > 0$ which as given in (3-3-5) necessarily follows from the assumption of concavity of (3-2-9). Thus it is possible that both conditions given in (3-3-20) and (3-3-21) are satisfied and hence that the effect of the introduction of new enforcement technology is to reduce the NPIT-optimal amount of resources devoted to enforcement activities and to raise the NPIT-optimal tax rate.

Denoting the right-hand sides of the conditions in (3-3-20) and (3-3-21) by $k_1$ and $k_2$ respectively, $dL_e^*/d\mu < 0$ and $d\Omega^*/d\mu > 0$ if and only if

$$k_2 < \left[[h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \mu} + \left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial \mu}\right] < k_1 \tag{3-3-23}$$

where $k_2 < k_1$ and $k_1, k_2 < 0$.

Examining the right-hand side inequality of (3-3-23) and rearranging gives

$$(3\text{-}3\text{-}24) \quad [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \mu} < k_1 - \left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial \mu}$$

Following the assumptions and results from Appendix 3-2 used above in the proof of this Proposition, the left-hand side of (3-3-24) is some positive number. Given that $k_1 < 0$, the inequality in (3-3-24) is satisfied only if $k_1$ is sufficiently small in absolute value which, from (3-3-20), corresponds to the assumption in the Proposition that the increase in marginal welfare impact of the tax rate, induced by the introduction of the new enforcement technology and denoted by S, is sufficiently small.

Rearranging (3-3-23) reveals that, given the assumptions of the Proposition, $dL_e^*/d\mu < 0$ and $d\Omega^*/d\mu > 0$ if and only if

$$(3\text{-}3\text{-}25) \quad \frac{k_1-\left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial \mu}}{[h(.)-SC_1(.)]} < \frac{\partial^2 Q}{\partial L_e \partial \mu}$$

$$< \frac{k_2-\left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial \mu}}{[h(.)-SC_1(.)]}$$

Thus, given that $\partial^2 Q/\partial L_e \partial \mu < 0$ by assumption in (A3-1-10), (3-3-25) shows that if the increase in marginal effectiveness of enforcement is bounded from above and, depending on the magnitude of $k_2$, from below, then the introduction of improved enforcement technology results in an increase in the NPIT-optimal rate of sales tax and a reduction in the NPIT-optimal level of resources devoted to enforcement.[5]

□

Proposition 3-3-3 deals with the introduction of improved enforcement technology and its results essentially reflect the substitution and output effects of this technical progress. It is clear that the introduction of technology which significantly enhances the marginal deterrent capabilities of both enforcement resources and the regulatory instrument could lead to a greater degree of control over industry output being optimal. This may occur as in result (iii) where more resources are devoted to enforcement activities or as in result (iv) where less resources are employed in enforcement but, as shown by the directional change in the tax rate, a greater level of restriction is achieved.

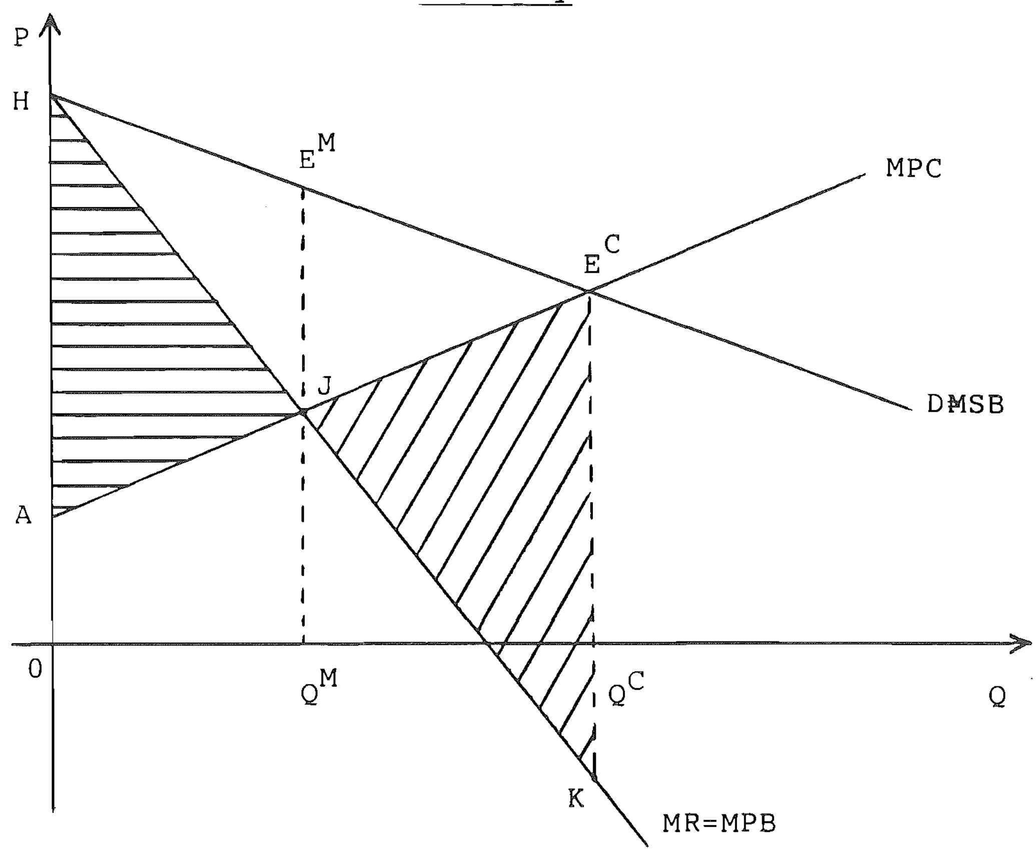
In the first case the substitution effect reinforces the output effect so that the increased marginal effectiveness of enforcement now outweighs its previously prohibitive marginal cost and hence it is optimal to devote more resources to enforcement. In the second case the output effect dominates the substitution effect. The assumption that $\partial Q/\partial \mu < 0$ implies that any level of enforcement activity is enhanced by the new technology resulting in greater deterrence and a reduced output level. The increase in marginal effectiveness of enforcement in this case is small relative to this output effect. The NPIT regulator then reduces the level of resources devoted to enforcement. Deterrence remains higher, and hence regulated equilibrium output lower, than before the introduction of the technology. This is implied by (3-2-11) given that the optimal tax rate rises.

The case in part (i) of the Proposition when both the optimal tax rate and optimal enforcement levels are reduced is not immediately obvious. Given that $\partial Q/\partial \mu < 0$, if the new technology did not alter the marginal deterrent capabilities of enforcement activity, the marginal benefit of enforcement would fall as a result of the smaller difference between the demand price and marginal social cost at the reduced output level. If then the new technology increases the marginal efficiency of enforcement by only a small amount, it is optimal for the NPIT regulator to reduce the amount of enforcement. A similar argument applies to the tax rate. Recalling from (3-2-11) that the optimal tax rate is set so that $\partial Q/\partial \Omega = 0$, it is possible in this case that the substitution effect outweighs the output effect and that the degree of control over industry output is lessened.

## 3-4 CAPTURE THEORY: OPTIMAL POLICY

In contrast to the public interest motivation of the NPIT regulator, the CT regulator serves the private interest of the industry that is regulated.[6] In the present partial equilibrium framework the CT regulator is assumed to act so as to maximize aggregate profits generated in the industry. This is equivalent to the problem of forming and maintaining a cartel in a competitive industry with a fixed number of firms. Figure 3-4-1 below illustrates the situation.

Figure 3-4-1: Comparison of profit levels under competitive behaviour and monopoly control of an industry.




The competitive industry equilibrium at $E^C$ with output level $Q^C$ involves a loss of potential industry profits of $JKE^C$. If the formation and operation of a cartel was costless, the cartel would operate as a pure monopolist and produce at output level $Q^M$ where marginal revenue, which shows the marginal private benefit of output (MPB), equals marginal private production cost and industry profits are maximized at AJH. As previously mentioned, $Q^M$ may or may not exceed the first-best output level $\hat{Q}$ in Figure 2-2-2, but, irrespective of this, it sets a lower bound on the optimal regulated output under CT regulation.

Given that any effective regulation requires enforcement and that any enforcement activity incurs strictly positive and increasing resource costs, it is not unreasonable to assume that the industry will be levied some fraction of the total cost of regulating itself. Using (3-2-2) the resource cost of enforcement incurred by the industry is

$$(3\text{-}4\text{-}1) \qquad b\omega(L_e) \quad ; \quad 0 \leq b \leq 1$$

In addition, the industry faces the payment of any fines levied as a result of the detection of illegal activities. To be consistent with the assumption of funded enforcement activities, it is assumed that any revenue raised from fine payments is redistributed in some way throughout the economy.

Expected fine payments are given from the expected penalty function. Following Sections 2-4 and 2-5 and using (3-2-5) the expected penalty in the sales tax case is given by

$$(3\text{-}4\text{-}2) \qquad EP_t = (1-a)G(Q(L_e,t,Z),X(L_e,t,Z),L_e) \ ;$$
$$G(Q(.),X(.),0) = 0, \quad 0 \leq a \leq 1$$

where $X(L_e,t,Z)$ is the optimal declared output level of the industry at given values of some parameters contained in the parameter vector Z, and 'a' is the share of fine revenue that is redistributed to the industry.

Differentiating (3-4-2) with respect to $L_e$

$$(3\text{-}4\text{-}3) \qquad \frac{\partial EP_t}{\partial L_e} = (1-a)[G_1\frac{\partial Q}{\partial L_e} + G_2\frac{\partial X}{\partial L_e} + G_3]; \quad G_1,\frac{\partial X}{\partial L_e},G_3 \geq 0;$$
$$G_2,\frac{\partial Q}{\partial L_e} \leq 0$$

Given the signs of the component terms of (3-4-3), which result from the analysis of Sections 2-4 and 2-5, the effect on total fine payments of an increase in resources devoted to enforcement in the case of a sales tax is ambiguous. The marginal fine revenue generated by additional enforcement, however, is assumed to be strictly decreasing at any level of enforcement for which the expected penalty is defined and hence $\partial^2 EP_t/\partial L_e^2 < 0$.

From (2-6-3) and (3-2-5) the expected penalty in the case of an output quota is given by

$$(3\text{-}4\text{-}4) \quad EP_R = (1-a)H(Q(L_e,R,Z),L_e,R); \; H(Q(.),0,R) = 0, \quad 0 \leq a \leq 1$$

where 'a' is as in (3-4-2) above. Differentiating,

$$(3\text{-}4\text{-}5) \quad \frac{\partial EP_R}{\partial L_e} = (1-a)[H_1\frac{\partial Q}{\partial L_e} + H_3]; \; H_1,H_3 \geq 0; \; \frac{\partial Q}{\partial L_e} \leq 0$$

and hence the effect on total fine payments of an increase in resources devoted to enforcement in the case of an output quota is also ambiguous. As in the case of a sales tax, $\partial^2 EP_R/\partial L_e^2 < 0$.

Therefore, using (3-4-1) and (3-4-5), the total cost of any level of enforcement activity incurred by the industry in the case of an output quota is

$$(3\text{-}4\text{-}6) \qquad w_R = b\omega(L_e) + EP_R$$

Differentiating with respect to $L_e$

$$(3\text{-}4\text{-}7) \qquad \frac{\partial w_R}{\partial L_e} = b\omega'(L_e) + \frac{\partial EP_R}{\partial L_e} > 0 \; ; \quad \frac{\partial^2 w_R}{\partial L_e^2} \geq 0$$

From (3-2-2), $\omega'(L_e) > 0$ and $\omega''(L_e) > 0$ and it is assumed here in (3-4-7) that the convexity of the resource cost of enforcement dominates the concavity of the expected penalty payments.

In the case of regulation by sales tax, there is the additional cost to the industry of the tax paid on any output that may be declared. This is given by

$$(3\text{-}4\text{-}8) \qquad \tau = \tau(X(L_e,t,Z),t)$$

Differentiating with respect to $L_e$

$$(3\text{-}4\text{-}9) \qquad \frac{\partial\tau}{\partial L_e} = \tau_1\frac{\partial X}{\partial L_e} \geq 0 \ ; \ \frac{\partial^2\tau}{\partial L_e^2} < 0$$

assuming that an increase in enforcement increases compliance with the tax at a diminishing rate; and with respect to the tax rate

$$(3\text{-}4\text{-}10) \qquad \frac{\partial\tau}{\partial t} = \tau_1\frac{\partial X}{\partial t} + \tau_2 \ ; \quad \tau_1,\tau_2 > 0 \ ; \ \frac{\partial X}{\partial t} < 0$$

From its component terms, (3-4-10) also has ambiguous sign. It is assumed that the reduction in compliance as the tax rate is increased at any given enforcement level produces tax revenues somewhat akin in shape to the Laffer curve and hence $\partial^2\tau/\partial t^2 < 0$.[7]

Using (3-4-1), (3-4-2) and (3-4-8), the total cost of any level of enforcement activity incurred by the industry in the case of a sales tax is

$$(3\text{-}4\text{-}11) \qquad w_t = b\omega(L_e) + EP_t + (1-a)\tau$$

Differentiating (3-4-11) with respect to the level of enforcement activity

$$(3\text{-}4\text{-}12) \qquad \frac{\partial w_t}{\partial L_e} = b\omega'(L_e) + \frac{\partial EP_t}{\partial L_e} + (1-a)\frac{\partial\tau}{\partial L_e} > 0; \ \frac{\partial^2 w_t}{\partial L_e^2} \geq 0$$

As in (3-4-7) it is assumed that the resource cost of enforcement is the dominant component in determining the properties of the net cost of enforcement to the industry.

A special case of the preceding analysis occurs when the funding of enforcement activities and dispersal of fine revenues is carried out on a per capita poll-tax basis. In this case, assuming that the industry comprises z individual agents, each of which owns one firm, from an economy of m agents,

$$0 < a = b = \frac{z}{m} < 1 \tag{3-4-13}$$

For enforcement of either regulatory instrument to take place, it is necessary that industry profits are increased through enforcement activities. Assuming that this is the case for initial enforcement efforts, and that the marginal benefit of enforcement is maximized at $Q^C$ and strictly decreases to zero at $Q^M$ while, from (3-4-7) and (3-4-12), the marginal cost of enforcement is non-decreasing, the optimal regulated equilibrium output level under CT regulation occurs at some intermediate output $\tilde{Q}$ where

$$\tilde{Q} = Q(\tilde{L}_e, \tilde{\Omega}, Z) \tag{3-4-14}$$

is the regulated equilibrium consistent with the CT optimal levels of enforcement ($\tilde{L}_e$) and the regulatory instrument ($\tilde{\Omega}$) at some fixed values of the parameters contained in the vector Z, such that

$$Q^M < \tilde{Q} < Q^C \tag{3-4-15}$$

The determination of $\tilde{Q}$ proceeds as follows, firstly for the case of an output quota.

(1) Regulation by Output Quota

Using (3-2-5), (3-2-7), (3-4-6), and the industry cost function which generates the supply curve of (2-2-7), the CT regulator chooses the size of quota and the amount of enforcement that maximizes profit within the regulatory framework according to the following expression.

$$(3\text{-}4\text{-}16) \qquad \underset{L_e, R}{\text{Maximize}}\ \pi_{R_e}(L_e,R,Y,\mu,a,b)$$

$$= h(Q(L_e,R,Y,\mu),Y)Q(L_e,R,Y,\mu)$$

$$- C(Q(L_e,R,Y,\mu)) - w_R(L_e,R,a,b)$$

Expression (3-4-16) shows the amount of increase in net profit as a result of reducing output from $Q^C$ in Figure 3-4-1 by enforced output quota and is assumed to be concave. The components of the parameter vector Z to be considered include respectively the level of aggregate consumer income (Y) and the state of enforcement technology ($\mu$). In addition, 'a' is the proportion of fine revenue that is redistributed to the industry and 'b' is the proportion of the resource cost of enforcement that is funded by the industry. These are assumed not to affect marginal production or consumption decisions and hence do not, of themselves, alter the regulated equilibrium.[8]

Differentiating (3-4-16) gives the following first-order conditions which describe a maximum.

$$(3\text{-}4\text{-}17) \qquad \frac{\partial \pi_{R_e}}{\partial L_e} = [h_1(Q(.),Y)Q(.) + h(Q(.),Y)$$

$$- C'(Q(.))]\frac{\partial Q}{\partial L_e} - \frac{\partial w_R}{\partial L_e} \leq 0 \ ; \ < \text{only if } \tilde{L}_e = 0$$

and

$$(3\text{-}4\text{-}18) \qquad \frac{\partial \pi_{R_e}}{\partial R} = [h_1(Q(.),Y)Q(.) + h(Q(.),Y)$$

$$- C'(Q(.))]\frac{\partial Q}{\partial R} - \frac{\partial w_R}{\partial R} \gtrless 0 \ ; \ > \text{only if } \tilde{R} = \tilde{Q}$$

$$< \text{only if } \tilde{R} = 0$$

When solved simultaneously, these conditions generate the optimal level of enforcement and size of the quota as functions of the various parameters, thus

$$(3\text{-}4\text{-}19) \qquad \tilde{L}_e = \tilde{L}_e(a,b,\mu,Y) \;;\; \tilde{R} = \tilde{R}(a,b,\mu,Y)$$

The initial bracketed term of (3-4-17) gives the difference between marginal revenue and marginal private cost which, when multiplied by the output effect of enforcement, gives the marginal benefit of enforcement. The second term gives the marginal cost of enforcement to the industry which, from (3-4-7), is positive and non-decreasing. At the optimum level of enforcement, assuming that enforcement takes place, these are equated. If the marginal cost of enforcement everywhere exceeds its marginal benefit, no enforcement will occur. The assumption that this is not the case is implicit in (3-4-15).

The component terms of (3-4-18) have a similar interpretation to those of (3-4-17) and show the net marginal benefit to the industry of increasing the quota. From (3-4-4) and (3-4-6)

$$(3\text{-}4\text{-}20) \quad \frac{\partial w_R}{\partial R} = \frac{\partial EP_R}{\partial R} = (1-a)\left[H_1\frac{\partial Q}{\partial R} + H_2\right] \;;\; H_1, \frac{\partial Q}{\partial R} > 0;\; H_2 < 0$$

Following Propositions 2-6-1 and 2-6-2 it is assumed that, as an increase in quota reduces the extent of illegal output and reduces the marginal expected penalty at every illegal output level, $H_2$ dominates the output effect of the quota and hence $\partial w_R/dR < 0$. If the marginal benefit of increasing the available quota everywhere exceeds the marginal cost of so doing, it is optimal for the CT regulator to set the quota at the regulated equilibrium output level of the industry.

Conditions (3-4-17) and (3-4-18) must be satisfied simultaneously at the CT-optimal regulated equilibrium output level $\tilde{Q}$.

LEMMA 3-4-1: Suppose an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax each of which is enforced by means of an expected monetary penalty. Assuming that an increase in enforcement reduces the equilibrium output level of the industry and that the marginal cost of enforcement to the industry is strictly positive, at the CT-optimal regulated industry equilibrium output level the marginal private production cost exceeds marginal revenue.

Proof.

In the case of an output quota this follows immediately (3-4-17). Given that, from (3-4-7) and (3-4-12), the marginal cost of enforcement is strictly positive and that, following Propositions 2-5-1 and 2-6-2, $\partial Q/\partial L_e < 0$ for both an output quota and sales tax provided that the expected penalty is directly binding on behaviour, then, if any enforcement is to occur under the CT framework so that $\tilde{L}_e > 0$, the reduction in industry output which enforcement induces must increase industry profitability. This necessitates that at the output level associated with $\tilde{L}_e$ marginal cost exceeds marginal revenue. That is

$$(3\text{-}4\text{-}21) \qquad h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.)) < 0$$

If however $\tilde{L}_e = 0$ then no deterrent to non-compliance with the regulatory instrument exists and hence the regulated equilibrium coincides with the unregulated competitive equilibrium. Recalling the assumption that the unregulated competitive equilibrium level in this industry exceeds the monopoly profit-maximizing level, (3-4-21) holds in this case also. Hence at any feasible CT-optimal regulated equilibrium output level, marginal private production cost exceeds marginal revenue.

□

<u>PROPOSITION 3-4-1:</u> <u>Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to a strictly binding output quota which is enforced by means of an expected monetary penalty. Given that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation and that the marginal cost of enforcement is non-decreasing, CT-optimal regulation of the industry exhibits the following characteristics:</u>

(i) <u>Given that the expected penalty is binding on the industry's</u>

behaviour and that enforcement increases industry profitability, the optimal amount of enforcement occurs where the marginal private benefit of enforcement equals its marginal private cost and is uniquely determined giving a regulated equlibrium output which exceeds the pure monopoly level.

(ii) Given that the optimal regulatory policy occurs when both first-order conditions are simultaneously satisfied, the optimal level of quota depends on the form and magnitude of the expected penalty. If the expected penalty function is strictly convex in the extent of constraint violation and does not include a flat rate component the optimal quota level occurs where the marginal private cost of increasing the quota equals the marginal private benefit of so doing and is uniquely determined at a level which is strictly less than the regulated equilibrium output of the industry. It is possible in these circumstances that the marginal cost to the industry of increasing the quota

<u>everywhere exceeds the marginal benefit of so doing and hence that the CT-optimal quota is zero.</u>

(iii) <u>If the quota is enforced by a constant unit-rate marginal expected penalty the CT-optimal quota is set at the regulated equilibrium output level that is consistent with $\tilde{L}_e$ from Part (i) of the Proposition.</u>

<u>Proof</u>.

(i) The assumptions that the expected penalty is binding on behaviour and that the firm's profit function is concave ensures, following Proposition 2-6-2, that $\partial Q/\partial L_e < 0$. The result then follows using (3-4-7) and Lemma 3-4-1.

(ii) Given that, from part (i) of the Proposition, it is optimal to control the industry so that $\tilde{L}_e > 0$, there is some level of quota strictly less than the unregulated competitive output level that is optimal for the CT regulator. This results from the fact that enforcement occurs only if $\partial Q/\partial L_e < 0$ which requires that there be some binding constraint to enforce.

Following Proposition 2-6-2, with an expected penalty function that is convex in the extent of constraint violation, $\partial Q/\partial R > 0$ and any increase in available quota that is binding on behaviour increases regulated equilibrium output. From Lemma 3-4-1 therefore, which shows that marginal cost exceeds marginal revenue over the relevant output range, an increase in quota of itself acts to reduce industry profits. This corresponds to the first composite term of (3-4-18) which thus represents the marginal cost to the industry of quota expansion. From (3-4-20), however, an increase in the quota reduces the expected penalty faced by the industry. This signs the second term of (3-4-18) which thus represents the marginal benefit of increasing the quota. The CT-optimal quota occurs where the marginal benefit and marginal cost of quota expansion are equated.

Given the assumptions that the expected penalty function is convex in the extent of constraint violation and contains no fixed unit rate component, the expected penalty is zero at the output level corresponding to full compliance with the quota and increases continuously as output expands illegally. With this form of expected penalty function, the marginal expected penalty for the initial illegal unit of output is less than the difference between marginal production cost and the demand price at the quota level of output. Hence, from the assumption of competitive profit-maximizing behaviour, full compliance with the quota is impossible to achieve through enforcement and therefore the CT-optimal output quota is strictly less than the regulated equilibrium output level consistent with $\tilde{L}_e$ from part (i) of the Proposition.

If the marginal cost of quota expansion, given the CT-optimal enforcement level, everywhere exceeds the marginal benefit of so doing, then from (3-4-18) any strictly positive output quota reduces industry profit and hence the CT-optimal output quota is zero.

(iii) If the quota is enforced by a constant unit rate expected penalty function then, following Proposition 2-6-2, any change in available quota which remains binding on behaviour will not affect the regulated equilibrium output of the industry and hence $\partial Q/\partial R = 0$. In this case, using (3-4-20), any increase in quota which remains binding increases industry profit and therefore, from (3-4-18), the CT-optimal quota is set at the regulated equilibrium industry output level consistent with the CT-optimal enforcement level $\tilde{L}_e$.

□

(2) Regulation by Sales Tax

With reference to (3-4-15) the determination of the CT-optimal regulated equilibrium in the case of regulation by an enforced unit rate sales tax proceeds as follows.

Using (3-2-5), (3-2-7), (3-4-11), and the industry cost function which generates the supply curve of (2-2-7), and modifying these formulations to include the parameterization used in the above discussion of an output quota, the CT regulator chooses the amount of enforcement and the rate of sales tax which together maximize profit within the regulatory framework according to the following expression:

$$(3\text{-}4\text{-}22) \qquad \underset{L_e,t}{\text{Maximize}}\ \pi_{R_e}(L_e,t,Y,\mu,a,b)$$

$$= h(Q(L_e,t,\mu,Y),Y)Q(L_e,t,\mu,Y)$$

$$- C(Q(L_e,t,\mu,Y)) - w_t(L_e,t,\mu,a,b)$$

Expression (3-4-22) shows the amount of increase in net profit as a result of reducing output from $Q^C$ in Figure 3-4-1 by the enforced sales tax and is assumed to be concave. Differentiating gives the following first-order conditions which describe a maximum.

$$(3\text{-}4\text{-}23) \qquad \frac{\partial \pi_{R_e}}{\partial L_e} = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial Q}{\partial L_e}$$

$$- \frac{\partial w_t}{\partial L_e} \leq 0 \quad ; \qquad < \text{ only if } \tilde{L}_e = 0$$

$$(3\text{-}4\text{-}24) \qquad \frac{\partial \pi_{R_e}}{\partial t} = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial Q}{\partial t}$$

$$- \frac{\partial w_t}{\partial t} \leq 0 \quad ; \qquad < \text{ only if } \tilde{t} = 0$$

The initial bracketed term of (3-4-23) gives the difference between marginal revenue and marginal private production cost which, when multiplied by the output effect of enforcement, gives the marginal benefit of enforcement. The second term gives the marginal cost to the industry of enforcement which, from (3-4-12), is positive and non-decreasing. At the optimal level of enforcement, assuming that enforcement takes place, these are equated. If the marginal cost of enforcement everywhere exceeds its marginal benefit, no enforcement takes place. The assumption that this is not the case is implicit in (3-4-15).

The component terms of (3-4-24) have a similar interpretation to those of (3-4-23) and show the net marginal benefit to the industry of increasing the tax rate. Using (3-4-10) and (3-4-11)

$$(3\text{-}4\text{-}25) \qquad \frac{\partial w_t}{\partial t} = \frac{\partial EP_t}{\partial t} + (1-a)[\tau_1\frac{\partial X}{\partial T} + \tau_2]$$

where from (3-4-2)

$$(3\text{-}4\text{-}25) \qquad \frac{\partial EP_t}{\partial t} = (1-a)[G_1\frac{\partial Q}{\partial t} + G_2\frac{\partial X}{\partial T}]\ ;\ G_1 > 0;\ G_2, \frac{\partial X}{\partial t}, \frac{\partial Q}{\partial t} < 0$$

Given the signs of the component terms of (3-4-26), which result from the analysis of Sections 2-4 and 2-5, the effect on total fine payments of an increase in the tax rate is ambiguous. It is assumed however that the effect of the reduction in declared output dominates that of the reduction in actual output and hence that $\partial EP_t/\partial t > 0$.

Substituting in (3-4-25), and using the analysis of (3-4-10), it is assumed that the positive effect on expected penalty payments outweighs any possible decrease in tax payments so that $\partial w_t/\partial t > 0$. It is also assumed that $\partial^2 w_t/\partial t^2 \geq 0$ so that the marginal cost to the industry of increasing the tax rate is non-decreasing.

These results hold only if the tax rate is directly binding on industry behaviour. Once the tax rate has reached the level above which no output is declared, further increases in the tax rate have no effect on the regulated equilibrium or on optimal declaration strategy and hence $\partial Q/\partial t = 0$ and $\partial w_t/\partial t = 0$. Conditions (3-4-23) and (3-4-24) must be satisfied simultaneously at the CT-optimal regulated equilibrium output level $\tilde{Q}$. Solving gives the CT-optimal enforcement level ($\tilde{L}_e$) and tax rate

($\tilde{t}$) as functions of the parameters specified in (3-4-22), thus

$$(3\text{-}4\text{-}27) \qquad \tilde{L}_e = \tilde{L}_e(a,b,\mu,Y) \quad ; \quad \tilde{t} = \tilde{t}(a,b,\mu,Y)$$

PROPOSITION 3-4-2: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to a strictly positive unit rate sales tax which is enforced by means of an expected monetary penalty. Suppose also that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation and that the marginal cost of enforcement is non-decreasing; CT-optimal regulation of the industry exhibits the following characteristics:

(i) Given that the expected penalty function is binding on the industry's behaviour and that enforcement increases industry profitability, the optimal amount of enforcement occurs where the marginal private benefit of enforcement equals its marginal private cost and is uniquely determined giving a regulated equilibrium output which exceeds the pure monopoly level.

(ii) Given that the optimal regulatory policy occurs when both first-order conditions are simultaneously satisfied, the optimal tax rate depends on the form and magnitude of the expected penalty function. If the expected penalty function is strictly convex in the extent of constraint violation, does not include a flat rate component, and the matrix of second-order partial derivatives of the expected penalty function with respect to actual and declared output is non-singular, then the optimal tax rate is set at a level such that, given the optimal enforcement level $\tilde{L}_e$ from part (i) of the Proposition, either no output is declared or some output is declared, but never where all output is declared.

(iii) Suppose that the expected penalty function is strictly convex in the extent of the violation and does not include a flat rate component but that the matrix of

<u>second-order partial derivatives of the expected penalty function with respect to actual and declared output is singular such that the effect on the marginal expected penalty with respect to declared output of an increase in declared output is equal and opposite to that of an increase in actual output and equal to the effect of a change in actual output on the marginal expected penalty with respect to actual output. The optimal tax rate is not less than the difference between the demand price and marginal private production cost at the equilibrium output level consistent with $\tilde{L}_e$ from part (i) of the Proposition given that no output is declared.</u>

<u>Proof</u>.

(i) The assumptions that the expected penalty is binding on behaviour and that the firm's profit function is concave ensures, following Proposition 2-5-1, that $\partial Q/\partial L_e < 0$. The result then follows using (3-4-12) and Lemma 3-4-1.

(ii) Given that, from part (i) of the Proposition, it is optimal to control the industry so that $\tilde{L}_e > 0$, there is necessarily some strictly positive tax rate that is optimal for the CT regulator. This results from the implication of (3-4-23) that enforcement occurs only if $\partial Q/\partial L_e < 0$ which requires that there be some constraint to enforce.

Following Proposition 2-5-1, with an expected penalty function that is convex in the extent of undeclared output, $\partial Q/\partial t < 0$ and any increase in the rate of tax which is binding on behaviour decreases regulated equilibrium output.

From (3-4-24) the CT-optimal tax rate occurs where the marginal benefit and marginal cost, to the industry, of increasing the tax rate, are equated. Given that the marginal benefit is strictly decreasing and the marginal cost is non-decreasing when the tax rate is binding on behaviour, it is not optimal for the CT regulator to indefinitely raise a binding tax rate.

Whether or not the tax is directly binding on behaviour at the CT-optimal regulated equilibrium output level depends on the simultaneous evaluation of (3-4-23) and (3-4-24). It may be the case that the relative net marginal benefits to the industry are such that it is optimal to enforce a sales tax to the point where output is partially declared. Here the optimal tax rate is $\bar{t}$ such that

$$(3\text{-}4\text{-}28) \qquad t = h(Q(\tilde{L}_e,\bar{t},\mu,Y),Y) - C'(Q(\tilde{L}_e,\bar{t},\mu,Y)) < G_Q(Q(\tilde{L}_e,\bar{t},\mu,Y),0,\tilde{L}_e,\mu)$$

In this case the marginal benefit of any tax less than $\bar{t}$ exceeds its marginal cost, given $\tilde{L}_e$, and the marginal benefit of any enforcement less than $\tilde{L}_e$ exceeds its marginal cost, given $\bar{t}$.

Alternatively, it may not be optimal to enforce a tax to the point where output is declared. In this case, any tax rate $t > \bar{\bar{t}}$ is not directly binding on behaviour and hence at any $\bar{\bar{t}} > t$, $\partial Q/\partial t = 0$ and $\partial w_t(.)/\partial t = 0$. In these circumstances, any tax rate $t > \bar{\bar{t}}$ is optimal for the CT regulator where

$$(3\text{-}4\text{-}29) \quad \bar{\bar{t}} = h(Q(\tilde{L}_e, \bar{\bar{t}}, \mu, Y), Y) - C'(Q(\tilde{L}_e, \bar{\bar{t}}, \mu, Y))$$
$$= G_Q(Q(\tilde{L}_e, \bar{\bar{t}}, \mu, Y), 0, \tilde{L}_e, \mu)$$

Given that there is no flat rate component in the expected penalty, it is not possible to ensure full declaration of output with any strictly positive tax rate and from (3-4-23) it is not optimal to enforce a zero-rate tax.

(iii) Following Propositions 2-4-4 and 2-5-1, if the matrix of second-order partial derivatives of the expected penalty function with respect to actual and declared output is singular, then for any tax rate $t < \bar{\bar{t}}$ as expressed in (3-4-29), $\partial Q/\partial L_e = 0$, and thus, from (3-4-23), the enforcement level $\tilde{L}_e$ cannot be optimal. Hence, in these circumstances, the CT-optimal tax rate $\tilde{t}$ can be no less than the difference between demand price and marginal private production cost at the regulated equilibrium output level associated with the marginal expected penalty with respect to output given that no output is declared.

□

In the case where the matrix of second-order partial derivatives of the expected penalty function is singular such that $G_{xx} = G_{qq} = -G_{xq}$, it follows from Propostion 2-5-1 that increasing enforcement beyond the level at which the marginal expected penalty with respect to output, given that no output is declared, is equal to the tax rate, has no effect on the regulated equilibrium output level of the industry. Suppose that, beginning from a regulated equilibrium where the marginal expected penalty with respect to output, given that no output is declared, is equal to the tax rate, enforcement is increased. Ceteris paribus, this results in an increase in the marginal expected penalty with respect to output at the regulated equilibrium output level. Given demand price and production cost the marginal unit of output is not profitable to produce.

In order to reduce the marginal expected penalty with respect to output firms can reduce output, increase the level of declared output, or implement some combination of the two. Given that $G_{qq} = - G_{qx}$, the effect on the marginal expected penalty with respect to output of a unit reduction in actual output is the same as that of a unit increase in declared output. The dynamics of the competitive equilibrium ensure that the response of the firm to the enforcement-induced increase in the marginal expected penalty will be to increase the amount of declared output at the margin as a decision to decrease actual output would create a divergence between the penalty-inclusive marginal cost of production and the demand price on the marginal unit of output.

The fact that the regulated equilibrium output level is not affected by an increase in enforcement which raises the marginal expected penalty with respect to output, given that no output is declared, beyond the level of the tax rate, implies from (3-4-23) that such a change in enforcement is not optimal. Conversely this implies that the CT-optimal tax rate is no less than the marginal expected penalty with respect to output, given that no output is declared, generated by the CT-optimal level of enforcement which, at the regulated equilibrium output level, is equal to the difference between marginal production cost and the demand price on the marginal unit of output as in part (iii) of the Proposition.

In the case where the expected penalty function is strictly convex in the extent of constraint violation, does not contain a flat-rate component, and the matrix of second-order partial derivatives of the expected penalty function with respect to actual and declared output is non-singular, as described in part (ii) of Proposition 2-5-1 an increase in enforcement in the circumstances described above acts to reduce the regulated equilibrium output level of the industry. Following (2-4-33), $G_{qq}G_{xx} - (G_{xq})^2 > 0$ in this instance and therefore it will not be the case that $G_{xx} = -G_{qx}$ or that $G_{qq} = -G_{xq}$. An increase in the amount of output that is declared in response to the increase in enforcement will then reduce the marginal expected penalty with respect to output but will not preserve the equality in absolute values between $G_x$ and $G_q$ that is necessary at equilibrium. If, however, it is the case that $G_{xx} = -G_{qx}$ then it also follows from (2-4-33) that $G_{qq} > -G_{qx}$. That is, the firm can reduce the marginal expected penalty with respect to output at a faster rate by reducing actual output than by increasing declared output. In either case, the optimal response of the firm involves a combination of an increase in declared output and a decrease in actual output. It is possible therefore, depending on the marginal benefit and marginal resource and tax liability cost of so doing, that it is optimal for a CT regulator to enforce partial compliance with a sales tax.

## 3-5 CAPTURE THEORY: POLICY RESPONSES TO PARAMETER CHANGES

Expressions (3-4-19) and (3-4-27) show that the CT regulator's optimal policy decisions are functions of various parameters. Assuming interior optimum solutions for all policy instruments, the response of the regulator to changes in parameter values can be examined. This is done here firstly for the case of regulation by output quota.

### (1) Regulation by Output Quota

Totally differentiating (3-4-17) gives

$$(3\text{-}5\text{-}1)\qquad A dL_e + B dR + E d\mu + F dY + J da + K db = 0$$

where

$$A = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e^2} - \frac{\partial^2 w_R}{\partial L_e^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

$$B = [h_1(Q(.),Y)Q(.)+ h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial R} - \frac{\partial^2 w_R}{\partial L_e \partial R}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial R}\right]\frac{\partial Q}{\partial L_e}$$

$$E = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu} - \frac{\partial^2 w_R}{\partial L_e \partial \mu}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial L_e}$$

$$F = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial Y} - \frac{\partial^2 w_R}{\partial L_e \partial Y}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}\right.$$

$$\left. + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)\right]\frac{\partial Q}{\partial L_e}$$

$$J = \frac{-\partial^2 w_R}{\partial L_e \partial a}$$

$$K = \frac{-\partial^2 w_R}{\partial L_e \partial b}$$

and totally differentiating (3-4-18) gives

(3-5-2) $\quad MdL_e + NdR + Sd\mu + TdY + Uda + Vdb = 0$

where

$$M = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial R \partial L_e} - \frac{\partial^2 w_R}{\partial R \partial L_e}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial R}$$

$$N = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial R^2} - \frac{\partial^2 w_R}{\partial R^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial R}\right]\frac{\partial Q}{\partial R}$$

$$S = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial R \partial \mu} - \frac{\partial^2 w_R}{\partial R \partial \mu}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial R}$$

$$T = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial R \partial Y} - \frac{\partial^2 w_R}{\partial R \partial Y}$$

$$+ \Big[ [h_{11}(Q(.),Y)Q(.)+h_1(Q(.),Y)+h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}$$

$$+ h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \Big] \frac{\partial Q}{\partial R}$$

$$U = -\frac{\partial^2 w_R}{\partial R \partial a}$$

$$V = -\frac{\partial^2 w_R}{\partial R \partial b}$$

Rewriting (3-5-1) and (3-5-2) in a matrix form and solving gives

$$(3\text{-}5\text{-}3) \quad \begin{bmatrix} dL_e \\ dR \end{bmatrix} = -\frac{1}{\Delta} \begin{bmatrix} N & -B \\ -M & A \end{bmatrix} \begin{bmatrix} Ed\mu + FdY + Jda + Kdb \\ Sd\mu + TdY + Uda + Vdb \end{bmatrix}$$

where $\begin{bmatrix} A & B \\ M & N \end{bmatrix}$ is the 2 x 2 matrix of coefficients of the terms in $dL_e$ and $dR$ from (3-5-1) and (3-5-2) and

$$(3\text{-}5\text{-}4) \qquad \Delta = AN - BM > 0$$

is the determinant of this coefficient matrix and is positive by the assumption of concavity of (3-4-16). Concavity also implies here that A and N are negative. The first parameter change considered concerns the level of aggregate consumer income.

(a) The level of aggregate consumer income

PROPOSITION 3-5-1: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to a strictly binding output quota enforced by means of an expected monetary penalty which is convex in output and the extent of quota violation. Suppose also that regulating the output of the industry necessitates the use of scarce resources for enforcement, that the CT regulator's objective function is strictly concave, and that the CT-optimal regulatory policy at some fixed vector of parameter values involves strictly positive levels of enforcement and quota in accordance with Proposition 3-4-1.

(i) Suppose that a change in aggregate consumer income leaves the marginal productivity of the regulator's policy instruments unchanged. An increase in aggregate consumer income, which increases demand for the output of the regulated commodity, reduces (increases) the CT-optimal level of enforcement and increases (reduces) the CT-optimal quota level if the combined effect of the income-induced increase in demand price and change in the price responsiveness of demand, evaluated at the initial regulated equilibrium output level, exceeds (is less than) the income-induced change in the per unit loss of industry profit on the marginal unit of output together with the change in the marginal cost to the industry of restricting output through the use of the relevant regulatory policy instrument. Ceteris paribus, the increase in income is more likely to lead to an increase in enforcement and a decrease in quota if it reduces the price responsiveness of demand than if it increases it.

(ii) Suppose that an increase in aggregate income, which increases demand for the output of the regulated industry, improves (reduces) the marginal productivity of the regulator's policy instruments. If the combined effect of the income-induced rise in demand price and change in price responsiveness of demand is less than (exceeds) the income-induced change in the per unit loss of industry profit on the marginal unit of output, the increase in income raises (lowers) the CT-optimal level of enforcement and reduces (increases) the CT-

<u>optimal quota level. If, however, the combined effect of the income-induced rise in demand price and change in the price responsiveness of demand exceeds (is less than) the income-induced change in the per unit loss of industry profit on the marginal unit of output, the effect of the increase in income on CT-optimal regulatory policy is ambiguous</u>.

<u>Proof</u>.

(i) Using (3-5-3) with $d\mu = da = db = 0$,

$$\frac{d\tilde{L}_e}{dY} = -\frac{N}{\Delta}.F + \frac{B}{\Delta}.T \tag{3-5-5}$$

and

$$\frac{d\tilde{R}}{dY} = \frac{M}{\Delta}.F - \frac{A}{\Delta}.T \tag{3-5-6}$$

Following Proposition 2-6-2, $\partial Q/\partial L_e < 0$ and $\partial Q/\partial R > 0$, while, from the assumption of concavity of (3-4-16), $\Delta > 0$ and $A,N < 0$. The discussion of the component terms of (3-5-3) in Appendix 3-3 shows in addition that $B,M < 0$. Using this information in (3-5-5) and (3-5-6) reveals that

$$\frac{d\tilde{L}_e}{dY} \gtrless 0 \text{ and } \frac{d\tilde{R}}{dY} \lesseqgtr 0 \tag{3-5-7}$$

$$\text{if } F \gtrless 0 \text{ and } T \lesseqgtr 0$$

Given the assumption that a change in aggregate consumer income does not affect the marginal productivity of the regulator's policy instruments, the component terms $\partial^2Q/\partial L_e\partial Y$ and $\partial^2Q/\partial R\partial Y$ from F and T, as defined in (3-5-1) and (3-5-2) respectively, are both zero. Following Appendix 3-1, an increase in income which increases demand for the output of the regulated commodity acts to increase the regulated equilibrium output of the industry and thus $\partial Q/\partial Y > 0$. From (3-5-1) then

$$F = \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\right]\frac{\partial Q}{\partial L_e} - \frac{\partial^2 w_R}{\partial L_e \partial Y} \tag{3-5-8}$$

Rearranging (3-5-7) $F \gtrless 0$ if and only if

$$h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \lesseqgtr \frac{\partial^2 w_R}{\partial L_e \partial Y}/\frac{\partial Q}{\partial L_e} - [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y} \tag{3-5-9}$$

Following (A3-3-10) $\partial^2 w_R/\partial L_e\partial Y < 0$ while from (A3-3-2) it is assumed that $[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))] < 0$. Given these results, and those of the previous discussion, the right-hand side of (3-5-9) is some positive number. The square bracketed term shows the unit rate of change in the difference between marginal revenue and marginal cost. Given that, from Lemma 3-4-1, marginal revenue is less than marginal cost over the relevant output range, multiplying this square bracketed term by $\partial Q/\partial Y$ gives the income-induced change in the unit loss of industry productive profit on the marginal unit of output. The other composite term on the right-hand side of (3-5-9) is the income-induced change in the marginal cost to the industry of a unit of enforcement divided by the marginal productivity of enforcement which shows the income-induced change in the marginal cost to the industry of restricting output through enforcement.

On the left-hand side of (3-5-9), the term $h_2(Q(.),Y)$ denotes the income-induced change in demand price at the initial regulated equilibrium quantity which, following the assumption of the Proposition, is positive. The term $h_1(Q(.),Y)$ represents the slope of the demand curve which, as previously defined, is negative. The effect of a change in income on the slope of the demand curve is given by $h_{12}(Q(.),Y)$. If this term is negative (positive), an increase in income increases (reduces) the absolute value of the slope of the demand curve or alternatively reduces (increases) the price responsiveness of demand.

From (3-5-2), using the assumption of part (i) of the Proposition

$$(3\text{-}5\text{-}10) \quad T = \Big[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\Big]\frac{\partial Q}{\partial R} - \frac{\partial^2 w_R}{\partial R\partial Y}$$

Rearranging (3-5-10), $T \lesseqgtr 0$ if and only if

$$(3\text{-}5\text{-}11) \quad h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \lesseqgtr \frac{\partial^2 w_R}{\partial R\partial Y}\Big/\frac{\partial Q}{\partial R} - [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}$$

Following (A3-3-13) $\partial^2 w_R/\partial R\partial Y > 0$ while from (A3-3-2) it is assumed that the square bracketed term on the right-hand side of (3-5-11) is positive. Given these results, and those of previous discussions, the right-hand side of (3-5-11) is some positive number. The interpretation of (3-5-11) is essentially the same as that of (3-5-9), the only difference being in the regulatory policy instrument which is relevant to each expression.

Using (3-5-9) and (3-5-11) in (3-5-7) reveals the validity of part (i) of the Proposition. From (3-5-7) the increase in income leads to an increase in CT-optimal enforcement and a decrease in the optimal quota level if $F > 0$ and $T < 0$. As shown in Appendix 3-3, these inequalities, from (3-5-9) and (3-5-11) respectively, are more likely to hold if $h_{12}(Q(.),Y) < 0$ than if $h_{12}(Q(.),Y) > 0$. Ceteris paribus, therefore, the increase in income is more likely to lead to an increase in CT-optimal enforcement and a decrease in the CT-optimal quota if it reduces the price responsiveness of demand than if it increases it.

(ii) Following (A3-3-10), if the change in income increases the marginal productivity of enforcement then $\partial^2 w_R/\partial L_e \partial Y$ is unambiguously negative. Using (3-5-9), if the combined effect of the income-induced rise in demand price and change in the price responsiveness of demand is less than the income-induced change in the per unit loss of industry profit on the marginal unit of output, the inequality is preserved by the addition of another positive term on the right-hand side of the expression. Using the discussion of (A3-3-13), in Appendix 3-3, a similar result can be derived from

(3-5-11). If, however, the change in income reduces the marginal productivity of the regulatory instruments then, following (A3-3-10) and (A3-3-13), the signs of $\partial^2 w_R/\partial L_e \partial Y$ and $\partial^2 w_R/\partial R \partial Y$ are ambiguous. It is possible in this case that the respective assumed inequalities in (3-5-9) and (3-5-11) are reversed. The effect of a change in income on CT-optimal regulatory policy in these circumstances is therefore ambiguous. The bracketed result of this part of the Proposition can be established by using the above argument with assumed inequalities reversed.

$\square$

Given that, following Lemma 3-4-1, marginal production cost exceeds marginal revenue over the relevant output range, the CT regulator increases industry profit by restricting output. An increase in income which raises the demand price at any given quantity may be expected to reduce the marginal benefit to the CT regulator of restricting output. The increase in income however also increases the regulated equilibrium output level of the industry at any given regulatory policy. Given the assumptions of a negatively sloped demand curve and strictly convex production cost, and the result of Lemma 3-4-1, the loss of aggregate surplus on the marginal unit of output is increased at the new equilibrium. This effect of itself would suggest an increase in the CT-optimal extent of control over the industry. Disregarding the effect of the change in income on the marginal cost to the industry of restricting output through the use of the respective regulatory policy instruments, the policy response of the CT regulator to the change in income depends on the relative size of these two effects together with the income-induced change in the price responsiveness of demand.

The Proposition showed that, ceteris paribus, an increase in income which reduces the price responsiveness of demand is more likely to lead to an increase in the CT-optimal extent of control over the industry than an increase in income that increases it. This is because the CT regulator is essentially acting as a monopolist in restricting the output of the industry and the less price responsive is demand, the greater is the potential profitability of exercising monopoly power.

Part (ii) of the Proposition considers the implications for CT regulatory policy if the increase in income affects the marginal productivity of the regulatory policy instruments. Depending on the magnitude and qualitative nature of these effects, the overall policy impact of a change in income may be either enhanced or reversed. For instance, an income-induced increase in the marginal productivity of enforcement leads to a ceteris paribus increase in the marginal benefit of additional enforcement to the CT regulator. This may cause CT-optimal enforcement to increase where, in the absence of such an effect, the reverse would hold.

The second parameter change considered concerns the proportion of the resource cost of enforcement that the industry is required to fund.

(b) <u>The resource cost of enforcement to the industry</u>

<u>PROPOSITION 3-5-2</u>: <u>Under the assumptions contained in the stem of Proposition 3-5-1, an increase in the proportion of the resource cost</u>

<u>of enforcement that the industry is required to fund reduces the extent to which it is optimal for the CT regulator to control the industry. Thus the CT-optimal level of enforcement activity is reduced and the CT-optimal quota size is increased.</u>

<u>Proof.</u>

Using (3-5-3) with $dY = da = d\mu = 0$ gives

(3-5-12) $$\frac{d\tilde{L}_e}{db} = -\frac{N}{\Delta}.K + \frac{B}{\Delta}.V$$

and

(3-5-13) $$\frac{d\tilde{R}}{db} = \frac{M}{\Delta}.K - \frac{A}{\Delta}.V$$

Following (3-4-20), $\partial^2 w_R/\partial R\partial b = 0$. Using (3-5-2) therefore, $V = 0$ and the signs of $d\tilde{L}_e/db$ and $d\tilde{R}/db$ depend only on the term in K. From (A3-3-6) and given the previously derived results on the signs of M, N and $\Delta$, $d\tilde{L}_e/db < 0$ and $d\tilde{R}/db > 0$. Hence an increase in the proportion of the resource cost of enforcement that the industry is required to fund reduces the CT-optimal enforcement level and increases the CT-optimal amount of quota. Given that, from Proposition 2-5-2, $\partial Q/\partial L_e < 0$, and that, from Proposition 2-6-1, $\partial Q/\partial R > 0$, these policy responses correspond to a reduction in the extent of control over the industry.

□

Proposition 3-5-2 is intuitively obvious. An increase in the requirement for the industry to fund the resource cost of enforcement unambiguously raises the marginal cost to the industry of restricting output and thus it is clearly profitable for the CT regulator to reduce the extent to which it controls the output of the industry.

The third parameter change considered concerns the proportion of fine payments that are redistributed to members of the industry.

(c) Net industry fine payments

PROPOSITION 3-5-3: Under the assumptions contained in the stem of Proposition 3-5-1 and given that the effect of an increase in available quota is to reduce the expected fine payments incurred by the industry, if an increase in enforcement activity increases the expected fine payments incurred by the industry then an increase in the proportion of fines that are redistributed to members of the industry increases the extent to which it is optimal for the CT regulator to control the industry. Thus the CT-optimal level of enforcement rises and the CT-optimal quota size is reduced. In these circumstances it is optimal for the CT regulator to unambiguously reduce the extent of control over the industry if an increase in enforcement activity sufficiently reduces the expected fine payments incurred by the industry.

Proof.

Using (3-5-3) with $dY = db = d\mu = 0$ gives

(3-5-14) $$\frac{d\tilde{L}_e}{da} = -\frac{N}{\Delta}.J + \frac{B}{\Delta}.U$$

and

(3-5-15) $$\frac{d\tilde{R}}{da} = \frac{M}{\Delta}.J - \frac{A}{\Delta}.U$$

From (3-4-20) the effect of an increase in available quota on the expected fine payments incurred by the industry, denoted as $\partial EP_R/\partial R$, is negative as assumed here if and only if

(3-5-16) $$[H_1\frac{\partial Q}{\partial R} + H_2] < 0$$

Differentiating (3-4-20) with respect to 'a'

(3-5-17) $$\frac{\partial^2 w_R}{\partial R\partial a} = -[H_1\frac{\partial Q}{\partial R} + H_2] > 0$$

which is positive by (3-5-16).

From (3-4-5), the effect of an increase in enforcement on the expected fine payments incurred by the industry $\partial EP_R/\partial L_e \gtrless 0$ if and only if

(3-5-18) $$[H_1\frac{\partial Q}{\partial L_e} + H_3] \gtrless 0$$

As shown in (A3-3-7)

(3-5-19) $$\frac{\partial^2 w_R}{\partial L_e\partial a} = -[H_1\frac{\partial Q}{\partial L_e} + H_3]$$

If, as assumed in the Proposition, $\partial EP_R/\partial L_e > 0$, then, from (3-5-18) and (3-5-19), $\partial^2 w_R/\partial L_e\partial a < 0$. Substituting this result and (3-5-17) into (3-5-14) and (3-5-15) then reveals that $d\tilde{L}_e/da > 0$ and $d\tilde{R}/da < 0$. Thus, if an increase in quota reduces expected fine payments incurred by the industry and an increase in enforcement increases them, it is optimal for the CT regulator to increase the extent to which it controls the industry in response to an increase in the proportion of refunded fine payments.

Control over the industry is unambiguously reduced only if enforcement is decreased given that the quota is not decreased or the quota is increased given that enforcement is not increased. If then $d\tilde{L}_e/da < 0$ and $d\tilde{R}/da > 0$, the extent of control over the industry is unambiguously decreased. Using (3-5-14) and (3-5-15)

$$(3\text{-}5\text{-}20) \qquad \frac{d\tilde{L}_e}{da} < 0 \text{ if and only if } \frac{\partial^2 w_R}{\partial L_e \partial a} > \frac{B}{N}\frac{\partial^2 w_R}{\partial R \partial a}$$

and

$$(3\text{-}5\text{-}21) \qquad \frac{d\tilde{R}}{da} > 0 \quad \text{if and only if } \frac{\partial^2 w_R}{\partial L_e \partial a} > \frac{A}{M}\frac{\partial^2 w_R}{\partial R \partial a}$$

where the right-hand side of the both conditions is positive and by (3-5-4)

$$(3\text{-}5\text{-}22) \qquad \frac{A}{M}\frac{\partial^2 w_R}{\partial R \partial a} > \frac{B}{N}\frac{\partial^2 w_R}{\partial R \partial a}$$

Following (3-5-20), (3-5-21) and (3-5-22), the extent of CT-optimal control over the industry is unambiguously reduced if the condition in (3-5-21) is satisfied. From (3-5-18) and (3-5-19) this requires that the expected fine payments incurred by the industry are sufficiently reduced by an increase in enforcement.

□

The interpretation of Proposition 3-5-3 is similar to that of Proposition 3-5-2. If the expected fine payments incurred by the industry increase with greater levels of enforcement and reductions in the amount of quota, an increase in the proportion of fine payments that are redistributed to members of the industry reduces the marginal cost to the CT regulator of increasing the extent to which output is restricted and allows this to profitably occur. The final parameter change considered concerns the state of enforcement technology.

(d) The state of enforcement technology

PROPOSITION 3-5-4: Under the assumptions contained in the stem of Proposition 3-5-1 the effect of an improvement in enforcement technology on CT-optimal regulatory policy depends on the relative sizes of its effect on the marginal productivity of the regulator's policy instruments and its effect on the regulated equilibrium at any given level of the policy instruments. If the new technology improves the marginal efficiency of enforcement and the quota sufficiently more than it reduces the regulated equilibrium output level at any policy combination, then it is optimal for the CT regulator to increase the extent to which it controls the industry. Thus the CT-optimal enforcement level is raised and the CT-optimal quota reduced. If, however, the effect of the new technology on the marginal efficiency of enforcement and the quota is sufficiently small relative to its effect on the regulated equilibrium output level at any policy combination, then it is optimal for the CT regulator to reduce enforcement and increase available quota.

Proof.

Using (3-5-3) with dY = da = db = 0 gives

(3-5-23) $$\frac{d\tilde{L}_e}{d\mu} = -\frac{N}{\Delta}.E + \frac{B}{\Delta}.S$$

and

(3-5-24) $$\frac{d\tilde{R}}{d\mu} = \frac{M}{\Delta}.E - \frac{A}{\Delta}.S$$

From (3-5-23) and (3-5-24) given that A,B,M,N < 0 and $\Delta > 0$ then $d\tilde{L}_e/d\mu > 0$ and $d\tilde{R}/d\mu < 0$ if E > 0 and F < 0 and $d\tilde{L}_e/d\mu < 0$ and $d\tilde{R}/d\mu > 0$ if E < 0 and F > 0. Following the discussion of (A3-3-15) and (A3-3-20), E > 0 if $\partial^2Q/\partial L_e\partial\mu$ is sufficiently negative which also requires that its absolute value be sufficiently large relative to that of $\partial Q/\partial\mu$, and F < 0 if $\partial^2Q/\partial R\partial\mu$ is sufficiently positive which requires that it be sufficiently large relative to the absolute value of $\partial Q/\partial\mu$.

Recalling the interpretation that $\partial^2Q/\partial L_e\partial\mu < 0$ and $\partial^2Q/\partial R\partial\mu > 0$ reflect the increase in marginal effectiveness of enforcement and the quota respectively as deterrents against illegal behaviour which is induced by the improvement in enforcement technology, and that $\partial Q/\partial\mu < 0$ shows the reduction in regulated equilibrium output at any combination of enforcement and quota caused by the improvement in technology, the result concerning the increase in the extent of CT-optimal control of the industry in response to the improvement in enforcement technology clearly follows.

Conversely, from (A3-3-15) and (A3-3-20), if

$-\partial^2Q/\partial L_e\partial\mu$ and $\partial^2Q/\partial R\partial\mu$ are sufficiently small in relation to $-\partial Q/\partial\mu$, $E < 0$ and $F > 0$ which, in (3-5-23) and (3-5-24), reveals that it is optimal for the CT regulator to reduce the level of enforcement and increase the amount of quota in response to the improvement in technology.

$\square$

Proposition 3-5-4 shows that the introduction of new enforcement technology which significantly enhances the marginal deterrent capabilities of both enforcement and the level of the quota can lead to a greater degree of control over the industry being optimal for the CT regulator. The case when it is optimal for the CT regulator to reduce the level of enforcement and increase the size of the quota has a similar interpretation to its NPIT counterpart in Proposition 3-3-3. Given that $\partial Q/\partial\mu < 0$, if the new technology did not alter the marginal deterrent capabilities of enforcement and the quota, the marginal benefit of both policy instruments would fall as a result of the smaller difference between marginal revenue and marginal cost at the reduced regulated equilibrium output level. If then the new technology increases the marginal efficiency of enforcement and the quota by only a small amount, it is optimal for the CT regulator to reduce the amount of enforcement and increase the size of available quota. Whether such action results in an overall increase or reduction in the extent of control over industry output depends on the relative sizes of these substitution and output effects of the introduction of the new technology.

This completes the analysis of CT-optimal regulation by output quotas. Regulation by taxes is now discussed.

(2) Regulation by Sales Tax

As shown in (3-4-27), the simultaneous solution of first-order conditions (3-4-23) and (3-4-24) yield CT-optimal values of the tax rate ($\tilde{t}$) enforcement level ($\tilde{L}_e$) as functions of the various parameters outlined in the previous analysis of CT regulation by output quota. Assuming interior optimum values for both policy instruments, their responses to changes in parameter values are now examined.

Totally differentiating (3-4-23) gives

$$(3\text{-}5\text{-}25) \quad AdL_e + Bdt + Ed\mu + FdY + Jda + Kdb = 0$$

where

$$A = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 0}{\partial L_e^{\,2}} - \frac{\partial^2 w_t}{\partial L_e^{\,2}}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

$$B = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial t} - \frac{\partial^2 w_t}{\partial L_e \partial t}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial t}\right]\frac{\partial Q}{\partial L_e}$$

$$E = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu} - \frac{\partial^2 w_t}{\partial L_e \partial \mu}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial L_e}$$

$$F = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial Y} - \frac{\partial^2 w_t}{\partial L_e \partial Y}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))\right]\frac{\partial Q}{\partial Y}$$

$$+ h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)]\frac{\partial Q}{\partial L_e}$$

$$J = -\frac{\partial^2 w_t}{\partial L_e \partial a}$$

$$K = -\frac{\partial^2 w_t}{\partial L_e \partial b}$$

and totally differentiating (3-4-24) gives

$$(3\text{-}5\text{-}26) \qquad MdL_e + Ndt + Sd\mu + TdY + Uda + Vdb = 0$$

where

$$M = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial t \partial L_e} - \frac{\partial^2 w_t}{\partial t \partial L_e}$$

$$+\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial t}$$

$$N = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial t^2} - \frac{\partial^2 w_t}{\partial t^2}$$

$$+\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial t}\right]\frac{\partial Q}{\partial t}$$

$$S = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial t \partial \mu} - \frac{\partial^2 w_t}{\partial t \partial \mu}$$

$$+\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial t}$$

$$T = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial t \partial Y} - \frac{\partial^2 w_t}{\partial t \partial Y}$$

$$+\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}\right.$$

$$\left. + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)\right]\frac{\partial Q}{\partial t}$$

$$U = -\frac{\partial^2 w_t}{\partial t \partial a}$$

$$V = -\frac{\partial^2 w_t}{\partial t \partial b}$$

Rewriting (3-5-24) and (3-5-25) in matrix form and solving gives

$$(3\text{-}5\text{-}27) \quad \begin{bmatrix} dL_e \\ dt \end{bmatrix} = -\frac{1}{\Delta} \begin{bmatrix} N & -B \\ -M & A \end{bmatrix} \begin{bmatrix} Ed\mu + FdY + Jda + Kdb \\ Sd\mu + TdY + Uda + Vdb \end{bmatrix}$$

where $\begin{bmatrix} A & B \\ M & N \end{bmatrix}$ is the 2 x 2 matrix of coefficients of the terms in $dL_e$ and dt from (3-5-25) and (3-5-26) and

$$(3\text{-}5\text{-}28) \qquad \Delta = AN - BM > 0$$

is the determinant of this coefficient matrix and is positive by the assumption of concavity of (3-4-22). Concavity also implies here that A and N are negative. The first parameter change to be considered concerns the level of aggregate consumer income.

(a) <u>The level of aggregate consumer income</u>.

<u>PROPOSITION 3-5-5</u>: <u>Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to a strictly positive unit rate sales tax which is enforced by means of an expected monetary penalty. Suppose also that the CT regulator's objective function is strictly concave and that CT-optimal regulatory policy at some fixed vector of parameter values involves strictly positive levels of enforcement and the tax rate in accordance with Proposition 3-4-2</u>.

(i) Suppose that a change in aggregate consumer income leaves the marginal productivity of the regulator's policy instruments unchanged. An increase in aggregate consumer income, which increases demand for the output of the regulated commodity, increases (reduces) the CT-optimal level of enforcement and the CT-optimal tax rate if the combined effect of the income-induced increase in demand price and change in the price responsiveness of demand, evaluated at the initial regulated equilibrium output level, is less than (exceeds) the income-induced change in the per unit loss of industry profit on the marginal unit of output together with the change in the marginal cost to the industry of restricting output through the use of the relevant policy instrument. Ceteris paribus, the increase in income is more likely to lead to an increase in enforcement and the tax rate if it reduces the price responsiveness of demand than if it increases the price responsiveness of demand. If the effect of the increase in income is to raise the marginal cost to the industry of restricting output through the use of regulatory policy, it is possible that it is optimal for the CT-regulator to increase the extent to which it controls the industry, in response to the change in income, only if the increase in income sufficiently reduces the price responsiveness of demand.

(ii) Suppose that an increase in aggregate income, which increases demand for the output of the regulated industry, improves (reduces) the marginal productivity of the regulator's policy instruments. If the combined effect of the income-induced rise in demand price and change in the price responsiveness of demand is less than (exceeds) the income-induced change in the per unit loss of industry profit on the marginal unit of output together with the change in the marginal cost to the industry of restricting output through the use of the relevant policy instrument, the increase in income raises (lowers) the CT-optimal level of

<u>enforcement and tax rate. If, however, the combined effect of the income-induced rise in demand price and change in the price responsiveness of demand exceeds (is less than) the income-induced change in the per unit loss of industry profit on the marginal unit of output together with the change in the marginal cost to the industry of restricting output through the use of the relevant policy instrument, the effect of the increase in income on CT-optimal regulatory policy is ambiguous.</u>

<u>Proof</u>.

(i) Using (3-5-27) with $d\mu = da = db = 0$

$$\frac{d\tilde{L}_e}{dY} = -\frac{N}{\Delta}.F + \frac{B}{\Delta}.T \tag{3-5-29}$$

and

$$\frac{d\tilde{t}}{dY} = \frac{M}{\Delta}.F - \frac{A}{\Delta}.T \tag{3-5-30}$$

Following Proposition 2-5-1, $\partial Q/\partial L_e$, $\partial Q/\partial t < 0$, while from the assumption of concavity of (3-4-22), $\Delta > 0$ and $A,N < 0$. The discussion of the component terms of (3-5-27) in Appendix 3-4 shows in addition that $B,M > 0$. Using this information in (3-5-29) and (3-5-30) reveals that

$$(3\text{-}5\text{-}31) \qquad \frac{d\tilde{L}_e}{dY} \text{ and } \frac{d\tilde{t}}{dY} \gtrless 0 \text{ if } F,T \lessgtr 0$$

Given the assumption that a change in aggregate consumer income does not affect the marginal productivity of the regulator's policy instruments, the component terms $\partial^2 Q/\partial L_e \partial Y$ and $\partial^2 Q/\partial t \partial Y$ from F and T, as defined in (3-5-25) and (3-5-26) respectively, are both zero. From (3-5-25) then

$$(3\text{-}5\text{-}32) \quad F = \Big[ [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \Big] \frac{\partial Q}{\partial L_e} - \frac{\partial^2 w_t}{\partial L_e \partial Y}$$

Rearranging (3-5-32), $F \gtrless 0$ if and only if

$$(3\text{-}5\text{-}33) \quad h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \lessgtr \frac{\partial^2 w_t}{\partial L_e \partial Y} / \frac{\partial Q}{\partial L_e} - [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}$$

The interpretation of the component terms of (3-5-33) is the same as that of their counterparts in (3-5-9). From (3-5-26), using the assumption of no productivity change

$$(3\text{-}5\text{-}34) \quad T = \Big[ [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y) \Big] \frac{\partial Q}{\partial t} - \frac{\partial^2 w_t}{\partial t \partial Y}$$

and rearranging, $T \gtrless 0$ if and only if

$$(3\text{-}5\text{-}35) \quad h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \lessgtr \frac{\partial^2 w_t}{\partial t \partial Y} / \frac{\partial Q}{\partial t} - [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}$$

As with (3-5-33), (3-5-35) has a counterpart in the proof of Proposition 3-5-1. The interpretation of the component terms of (3-5-35) is therefore the same as those of (3-5-11). Given these similarities, using (3-5-33) and (3-5-35) in (3-5-31), the validity of the first part of section (i) of Proposition 3-5-5 clearly follows.

If, following Appendix 3-4, conditions are such to ensure that $\partial^2 w_t/\partial L_e \partial Y$ and $\partial^2 w_t/\partial t \partial Y$ are both negative, the right-hand sides of (3-5-33) and (3-5-35) are unambiguously positive. Given the assumption in the Proposition that the increase in income raises the demand price of the regulated commodity at every quantity, $h_2(Q(.),Y) > 0$. Following the discussion of component terms in the proof of Proposition 3-5-1, $h_{12}(Q(.),Y) < 0$ reflects an income-induced reduction in the price responsiveness of demand. From (3-5-33) and (3-5-35), F and T are more likely to be positive if $h_{12}(Q(.),Y) < 0$ than if it is positive. Using (3-5-31) therefore, it follows that, ceteris paribus, the increase in income is more likely to lead to an increase in enforcement and the tax rate if it reduces the price responsiveness of demand than if it increases it.

If the effect of the increase in income is to raise the marginal cost to the industry of restricting output through the use of regulatory policy, $\partial^2 w_t/\partial L_e \partial Y$ and $\partial^2 w_t/\partial t \partial Y$ are both negative. The right-hand sides of (3-5-33) and (3-5-35) in this case are ambiguous in sign. It is possible in particular that the right-hand sides of both expressions are negative. If this is so it will be optimal for the CT regulator to increase the extent to which it controls the industry in response to the increase in income only if the increase in income sufficiently reduces the price responsiveness of demand.

(ii) If the increase in income acts to change the marginal productivity of the regulator's policy instruments, the terms $\partial^2Q/\partial L_e\partial Y$, $\partial^2Q/\partial t\partial Y$, $\partial^2X/\partial L_e\partial Y$ and $\partial^2X/\partial t\partial Y$ are non-zero. In particular, if the increase in income raises the marginal productivity of the regulatory instruments, $\partial^2Q/\partial L_e\partial Y$, $\partial^2Q/\partial t\partial Y < 0$ and $\partial^2X/\partial L_e\partial Y$, $\partial^2X/\partial t\partial Y > 0$. Incorporation of these terms in (A3-4-10) and (A3-4-12) reveals that the signs of $\partial^2w_t/\partial L_e\partial Y$ and $\partial^2w_t/\partial t\partial Y$ are ambiguous. The statement of the Proposition, however, circumvents this difficulty by assuming a clearly defined inequality between the combined effect of the income-induced rise in demand price and change in the price responsiveness of demand, and the income-induced change in the per unit loss of industry profit on the marginal unit of output together with the change in marginal cost to the industry of restricting output through the use of the relevant policy instrument. Assuming that the "less-than" inequality holds between these component terms and that the change in income leaves the marginal productivity of the regulatory instruments unchanged, then, from (3-5-33)

$$(3\text{-}5\text{-}36) \quad h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) < \frac{\partial^2 w_t}{\partial L_e \partial Y} / \frac{\partial Q}{\partial L_e}$$

$$- [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial Y}$$

Using the formulation of F from (3-5-25), given Lemma 3-4-1, the incorporation of the assumption concerning an income-induced rise in the marginal productivity of the regulatory instruments results in the addition of another positive term on the right-hand side of (3-5-36). The inequality in (3-5-36) is therefore preserved by this type of change. If, however, the effect of the change in income is to reduce the marginal productivity of the regulatory instruments, a negative term is added to the right-hand side of (3-5-36) and the inequality is not necessarily preserved. This argument can also be determined through the use of (3-5-35) and can be extended to cover the bracketed case by reversing assumed inequalities. Using (3-5-31) therefore, the validity of part (ii) of the Proposition is established.

□

The results and interpretation of Proposition 3-5-5 are essentially the same as those of Proposition 3-5-1 except that in this case increases in the level of enforcement and in the tax rate are mutually reinforcing deterrents. The second parameter change considered concerns the proportion of the resource cost of enforcement that the industry is required to fund.

(b) <u>The resource cost of enforcement to the industry</u>.

<u>PROPOSITION 3-5-6</u>: <u>Under the assumptions contained in the stem of Proposition 3-5-5, an increase in the proportion of the resource cost of enforcement that the industry is required to fund reduces the CT-optimal level of enforcement and the CT-optimal tax rate thus reducing the CT-optimal extent of control over industry output</u>.

<u>Proof</u>.

Using (3-5-27) with $d\mu = da = dY = 0$ gives

$$\frac{d\tilde{L}_e}{db} = -\frac{N}{\Delta}.K + \frac{B}{\Delta}.V \tag{3-5-37}$$

and

$$\frac{d\tilde{t}}{db} = \frac{M}{\Delta}.K - \frac{A}{\Delta}.V \tag{3-5-38}$$

From (3-4-25) and (3-4-26), as defined in (3-5-26), $V = 0$. Given that $\Delta > 0$ and $A, N < 0$ by the assumption of concavity of (3-4-22) and that, following the discussion in Appendix 3-4, $B, M > 0$, the signs of (3-5-37) and (3-5-38) depend only on the sign of K. In particular

$$\frac{d\tilde{L}_e}{db}, \frac{d\tilde{t}}{db} \gtrless 0 \text{ if and only if } K \lessgtr 0 \tag{3-5-39}$$

From (A3-4-5) and (3-5-25), $K < 0$ and hence, following (3-5-39) $d\tilde{L}_e/db$ and $d\tilde{t}/db$ are both negative.

□

Proposition 3-5-6, as was the case for Proposition 3-5-2, is intuitively obvious. The next parameter change considered concerns the proportion of fine payments that are redistributed to members of the industry.

(c) Net industry fine payments

PROPOSITION 3-5-7: Under the assumptions contained in the stem of Proposition 3-5-5, given that the effect of an increase in enforcement is to increase the overall value of expected fine payments and taxes that the industry incurs, if an increase in the tax rate similarly raises the overall value of fine payments and taxes which the industry incurs, then an increase in the proportion of fines and tax revenue that is redistributed to members of the industry increases the extent to which it is optimal for the CT regulator to control the industry. Thus the CT-optimal levels of enforcement and tax rate are raised. In these circumstances it is optimal for the CT regulator to unambiguously reduce the extent of control over the industry if an increase in the tax rate sufficiently reduces the overall value of expected fine payments and taxes incurred by the industry.

Proof.

Using (3-5-27) with $d\mu = db = dY = 0$ gives

(3-5-40) $$\frac{d\tilde{L}_e}{da} = -\frac{N}{\Delta}.J + \frac{B}{\Delta}.U$$

and

(3-5-41) $$\frac{d\tilde{t}}{da} = \frac{M}{\Delta}.J - \frac{A}{\Delta}.U$$

From (A3-4-6), the assumption of the Proposition here that $[\frac{\partial EP_t}{\partial L_e} + \frac{\partial \tau}{\partial L_e}] > 0$ ensures, assuming (3-5-25), that $J > 0$. Substituting into (3-5-40) and (3-5-41) reveals that $d\tilde{L}_e/da$ and $d\tilde{t}/da$ are both positive if $U > 0$. Using (A3-4-7) and (3-5-26), $U > 0$ if $[\frac{\partial EP_t}{\partial t} + \frac{\partial \tau}{\partial t}] > 0$. Hence, given that the overall value of expected fine payments and taxes rises as a result of increased enforcement, if an increase in the tax rate has the same qualitative effect, then an increase in the proportion of fine and tax revenues that are refunded to members of the industry has the effect of increasing the CT-optimal tax rate and level of enforcement.

Control over the industry is unambiguously reduced only if enforcement activity is reduced given that the tax rate is not increased or the tax rate is reduced given that enforcement is not increased. If then $d\tilde{L}_e/da$ and $d\tilde{t}/da$ are both negative, the extent of control over the industry by the CT regulator is unambiguously reduced. Using (3-5-40) and (3-5-41)

(3-5-42) $$\frac{d\tilde{L}_e}{da} < 0 \text{ if and only if } U < \frac{N}{B}.J$$

and

(3-5-43) $$\frac{d\tilde{t}}{da} < 0 \text{ if and only if } U < \frac{M}{A}.J$$

where, using (3-5-25) given the assumption that $\partial^2 w_t/\partial L_e \partial a < 0$, the right-hand side of both conditions is negative and, by concavity of (3-4-22)

(3-5-44) $$\frac{N}{B}.J < \frac{M}{A}.J$$

Following (3-5-42), (3-5-43) and (3-5-44), the extent of CT-optimal control over the industry is unambiguously reduced if the condition in (3-5-42) is satisfied. From (A3-4-7) this requires that an increase in the tax rate sufficiently reduces the overall value of expected fine payments and taxes incurred by the industry.

□

In comparison with Proposition 3-5-3, Proposition 3-5-7 is further complicated by the fact that in the case of regulation by a sales tax, changes in the rate of tax directly affect the cost of regulation to the industry. The assumption that the overall value of expected fines and taxes paid by the industry increases with increasing levels of enforcement and higher tax rates ensures that an increase in the proportion of these payments that are refunded to members of the industry reduces the marginal cost to the industry of increasing the extent to which output is restricted and allows this to profitably occur.

The final parameter change considered in this section concerns the state of enforcement technology.

(d) <u>The state of enforcement technology</u>.

<u>PROPOSITION 3-5-8</u>: <u>Under the assumptions contained in the stem of Proposition 3-5-1, the effect of an improvement in enforcement technology on CT regulatory policy depends on the relative sizes of its effect on the marginal productivity of the regulator's policy instruments and its effect on the regulated equilibrium at any given level of the policy instruments.</u>

(i) <u>If the new technology improves the marginal efficiency of enforcement and the tax rate sufficiently more than it reduces the regulated equilibrium output level at any given policy combination, then it is optimal for the CT regulator to increase the extent to which it controls the industry. Thus the CT-optimal enforcement level is raised and the CT-optimal tax rate increased. If, however, the effect of the new technology on the marginal efficiency of enforcement and the tax rate is sufficiently small relative to its effect on the regulated equilibrium output level at any policy combination, then it is optimal for the CT regulator to reduce both enforcement activity and the tax rate.</u>

(ii) <u>Given that the introduction of improved enforcement technology raises the marginal effectiveness of the tax rate as a deterrent sufficiently in relation to its effect on the regulated equilibrium output level so that the impact of a change in the tax rate on industry</u>

profits is increased, and that this increase in marginal profit impact is sufficiently small, it is possible that the effect of the new technology is to increase the CT-optimal tax rate and decrease the CT-optimal level of resources devoted to enforcement activity.

Proof.

(i) Using (3-5-27) with $dY = da = db = 0$ gives

(3-5-45) $$\frac{d\tilde{L}_e}{d\mu} = -\frac{N}{\Delta}.E + \frac{B}{\Delta}.S$$

and

(3-5-46) $$\frac{d\tilde{t}}{d\mu} = \frac{M}{\Delta}.E - \frac{A}{\Delta}.S$$

where E and S are the coefficients on the state of technology parameter in (3-5-25) and (3-5-26) respectively.

From (3-5-45) and (3-5-46), given that $\Delta > 0$ and $A,N < 0$ by the concavity of (3-4-22) and that $B,M > 0$ following the discussion of Appendix 3-4, then $d\tilde{L}_e/d\mu > 0$ and $d\tilde{t}/d\mu > 0$ if $E,S > 0$, and $d\tilde{L}_e/d\mu < 0$ and $d\tilde{t}/d\mu < 0$ if $E,S < 0$. Following the discussion of (A3-4-14) and (A3-4-17), $E,S > 0$ if $\partial^2Q/\partial L_e\partial\mu$ and $\partial^2Q/\partial t\partial\mu$ respectively are sufficiently negative which also requires that their absolute values be sufficiently large relative to that of $\partial Q/\partial\mu$.

Recalling the interpretation that $\partial^2Q/\partial L_e\partial\mu < 0$ and $\partial^2Q/\partial t\partial\mu < 0$ reflect the increase in marginal effectiveness, as deterrents against illegal behaviour, of enforcement and the tax rate respectively which is induced by the improvement in enforcement technology, and that $\partial Q/\partial\mu$ shows the reduction in regulated equilibrium output at any given policy combination which is caused by the improved technology, the result concerning the increase in CT-optimal tax rate and level of enforcement in response to the improvement in enforcement technology clearly follows.

Conversely, from (A3-4-14) and (A3-4-17), if $-\partial^2Q/\partial L_e\partial\mu$ and $-\partial^2Q/\partial t\partial\mu$ are sufficiently small in relation to $-\partial Q/\partial\mu$, E and S are negative which in (3-4-44) and (3-4-45) reveals that it is optimal for the CT regulator to reduce both the level of enforcement and the tax rate in response to the improvement in technology.

(ii) The assumption contained in the Proposition is that the effect of the new enforcement technology is to increase the marginal efficiency of the tax rate sufficiently such that, from (A3-4-17), $S > 0$. Using this and rearranging (3-4-44) and (3-4-45) gives

$$\frac{d\tilde{L}_e}{d\mu} < 0 \text{ if and only if } E < \frac{B}{N}.S \tag{3-5-47}$$

and

$$\frac{d\tilde{t}}{d\mu} > 0 \text{ if and only if } E > \frac{A}{M}.S \tag{3-5-48}$$

Given that the right-hand side of each condition is negative, both conditions can be simultaneously satisfied only if

$$(3\text{-}5\text{-}49) \qquad \frac{A}{M}.S < \frac{B}{N}.S$$

Rearranging, (3-5-49) simplifies to AN - BM > 0 which necessarily follows from the assumption of concavity of (3-4-22). Thus it is possible that both conditions given in (3-5-45) and (3-5-46) are satisfied and hence that the effect of the introduction of new enforcement technology is to reduce the CT-optimal amount of resources devoted to enforcement but increase the CT-optimal tax rate.

Denoting the right-hand sides of the conditions in (3-5-45) and (3-5-46) by $k_1$ and $k_2$ respectively and substituting for E from (3-5-25), $d\tilde{L}_e/d\mu < 0$ and $d\tilde{t}/d\mu > 0$ if and only if

$$(3\text{-}5\text{-}50) \quad k_2 < \left[ [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C''(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu} - \frac{\partial^2 w_t}{\partial L_e \partial \mu} + \left[ [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e} \right] \frac{\partial Q}{\partial \mu} \right] < k_1$$

where $k_2 < k_1$ and $k_1, k_2 < 0$.

Examining the right-hand side inequality of (3-5-50) shows that it holds if and only if

$$(3\text{-}5\text{-}51) \quad [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu} < k_1 + \left[ \frac{\partial^2 w_t}{\partial L_e \partial \mu} - \left[ [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y) - c''(Q(.))]\frac{\partial Q}{\partial L_e} \right] \frac{\partial Q}{\partial \mu} \right]$$

Following the assumptions and results in Appendix 3-4, the left-hand side of (3-5-51) is some positive number while the large bracketed term on the right-hand side, using (A3-4-14), is positive also. Given that $k_1 < 0$, the inequality in (3-5-51) is satisfied only if the absolute value of $k_1$ is sufficiently small which, from the condition in (3-5-47), corresponds to the assumption in the Proposition that the increase in the marginal welfare impact of the tax rate, induced by the introduction of the new enforcement technology and denoted by S, is sufficiently small.

Rearranging (3-5-50) reveals that, given the assumptions contained in the Proposition, $d\tilde{L}_e/d\mu < 0$ and $d\tilde{t}/d\mu > 0$ if and only if

$$(3\text{-}5\text{-}52)\quad \frac{k_1+\frac{\partial^2 w_t}{\partial L_e \partial\mu}-\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial\mu}}{[h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]}$$

$$< \frac{\partial^2 Q}{\partial L_e \partial\mu}$$

$$< \frac{k_2+\frac{\partial^2 w_t}{\partial L_e \partial\mu}-\left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial\mu}}{[h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C'(Q(.))]}$$

Thus, given that $\partial^2 Q/\partial L_e \partial\mu < 0$ by assumption, (3-5-51) shows that if the increase in marginal effectiveness of enforcement, which is induced by the introduction of the new technology, is bounded from above and, depending on the magnitude of $k_2$, from below, then the introduction of improved enforcement technology results in an increase in the CT-optimal tax rate and a reduction in the CT-optimal level of resources devoted to enforcement.

$\square$

The interpretation of result (i) of Proposition 3-5-8 is again similar to its counterpart, Proposition 3-5-4. Given that $\partial Q/\partial\mu < 0$, the introduction of the new technology reduces the marginal benefit of both policy instruments at any given policy combination unless it significantly improves their marginal deterrent capabilities. If their marginal deterrent capabilities are not sufficiently enhanced it will be optimal for the CT regulator to reduce both the tax rate and the level of enforcement. Whether or not this results in a reduction in the extent to which the output of the industry is restricted depends on the relative sizes of these substitution and output effects of the technical progress.

Finally result (ii) corresponds to result (ii) of Proposition 3-3-3 which concerns NPIT regulation by unit rate sales tax. The assumption that $\partial Q/\partial\mu < 0$ implies that any level of enforcement activity is enhanced by the new technology resulting in greater deterrence and a reduced output level. The increase in marginal effectiveness of enforcement in this case is small relative to this output effect. The CT regulator then reduces the level of resources devoted to enforcement. It is likely, however, that deterrence remains higher and hence regulated equilibrium output lower than before the introduction of the new technology given that the CT-optimal tax rate rises.

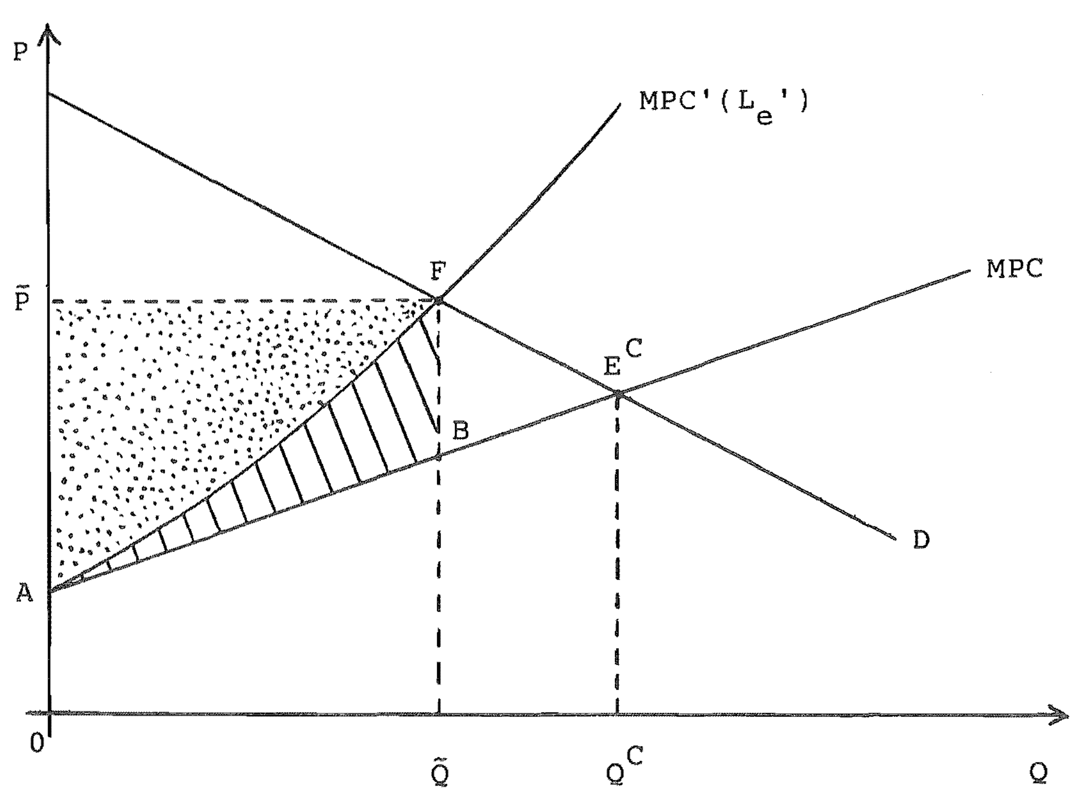
## 3-6 CAPTURE THEORY: THE CHOICE OF REGULATORY INSTRUMENT

From the analysis of Sections 2-4, 2-5 and 2-6, it is evident that the regulated equilibrium in a competitive industry occurs where the marginal expected penalty with respect to output is equal to the difference between the demand price and marginal private cost of production. Enforcement of the regulation generates this regulated equilibrium at a restricted industry output level by inflating the penalty-inclusive marginal cost structure of the firm and hence the industry.

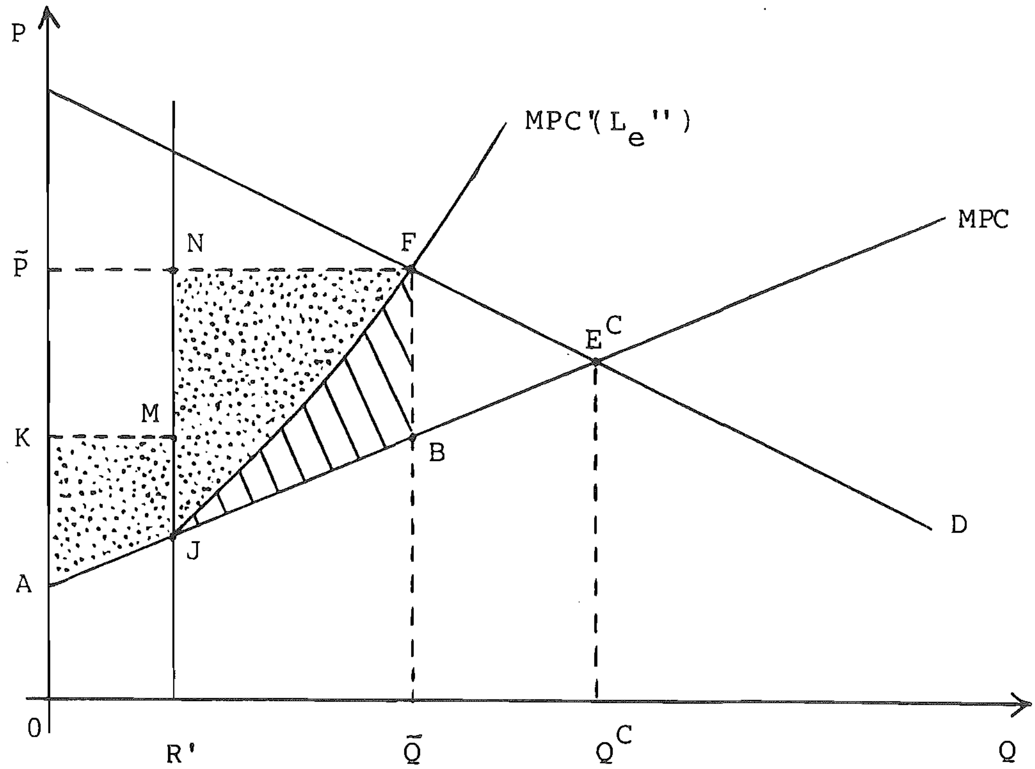
The CT regulator is motivated by the profitability of one sector of the economy only and thus will be interested in the distributional aspects of regulatory policy which will, in general, differ between the sales tax and output quota. This is illustrated in Figure 3-6-1 below.

Figure 3-6-1: A comparison of industry profitability under regulation by sales tax and output quota

(i) Zero output quota or sales tax with zero output declaration



(ii) Non-zero binding output quota or sales tax with partially declared output




Suppose there is some CT-optimal price/quantity combination ($\tilde{P}$,$\tilde{Q}$) which corresponds to a regulated industry equilibrium at F. Figure 3-6-1 shows several different policy combinations which generate regulated equilibrium at F. In each case MPC shows the marginal private cost of production and MPC' represents the penalty-inclusive marginal production cost. The form of the expected penalty function is fixed and given but the marginal expected penalty at any illegal output level varies with the amount of enforcement activity. It is assumed in constructing the diagram that the expected penalty function is convex in the extent of constraint violation and that the marginal expected penalty with respect to output at any given level of illegal behaviour is the same whether the regulatory instrument be a sales tax or an output quota.

Panel (i) of Figure 3-6-1 illustrates industry profitability in the case of either a zero output quota or a sales tax set at such a level so that no output is declared. Following the analysis of Sections 2-4 and 2-5 this requires a sales tax with unit rate at least as great as the marginal expected penalty at the regulated equilibrium output level, given that no output is declared, which corresponds to FB on the diagram. The preceding analysis of NPIT regulation revealed that the aggregate outcome of these two regulatory policies is identical and panel (i) of Figure 3-6-1 shows that the distributional impacts of the policies are identical also.

Given the assumptions which underly the diagram in panel (i) of Figure 3-6-1, area ABF$\tilde{P}$ shows the maximum potential value of monopoly profits to be realised by restricting industry output to $\tilde{Q}$. From this,however, a certain proportion, (1-a) using (3-4-2) or (3-4-4), of area ABF, which represents expected penalty payments incurred by the industry, must be deducted as must a proportion, 'b' using (3-4-1), of the resource cost of enforcement associated with the level of enforcement $L_e'$ which generates expected penalties consistent with MPC. This gives some total level of industry profitability $\pi_{R_e}'$ which is the same for each of the above-mentioned policies.

Panel (ii) of Figure 3-6-1 illustrates industry profitability at the same regulated equilibrium as in panel (i) but with a strictly positive and binding output quota or a sales tax with partially declared output. It is assumed in the diagram that both the output quota and the level of declared output are set at some size R'. The regulated industry equilibrium occurs at F because the marginal expected penalty with respect to output, given quota size or declared output level R', is again equal to FB, the difference between the demand price and marginal private cost of production at output level $\tilde{Q}$.

Given that the regulated equilibrium is unchanged between panel (i) and panel (ii) the value of potential monopoly profits at the restricted output level $\tilde{Q}_{R_e}$ shown by ABF$\tilde{P}$ is identical to that in panel (i). The total size of expected penalty payments incurred by the industry, JBF, is unambiguously smaller than its counterpart ABF in panel (i) and hence the residual area AJF$\tilde{P}$ exceeds AF$\tilde{P}$ in panel (i).

Industry profitability however is not unambiguously greater in the situation depicted in panel (ii) than it is in panel (i). Following Sections 2-4, 2-5 and 2-6, it is assumed that an increase in the marginal expected penalty with respect to output at any level of illegal behaviour requires an increase in the amount of enforcement activity. Hence the level of enforcement $L_e''$ associated with the marginal expected penalty with respect to output, FB, given quota or declaration level R' in panel (ii) exceeds that of $L_e'$ associated with marginal expected penalty with respect to output, FB, given a quota or declaration level of zero in panel (i). An evaluation of the relative profitability in the two situations then requires firstly a comparison of the saving in expected penalty payments and the extra costs incurred from the additional enforcement activity. Thus, denoting industry profitability in panel (ii) of Figure 3-6-1 as $\pi_{R_e}''$, a first approximation to the change in industry profitability between the two situations is, using (3-4-1), (3-6-1) $\quad \pi_{R_e''} - \pi_{R_e'} = (1-a)[ABF-JBF] - b[\omega(L_e'')-\omega(L_e')]$

Assuming that (3-6-1) is positive, it would appear that the CT regulator would prefer the situation in panel (ii) to that in panel (i). Industry profitability in this instance however differs between the two regulatory instruments. In the case of an output quota, the analysis of Section 2-6 showed that the unit trading price of the quota is the marginal expected penalty with respect to output evaluated at the regulated industry equilibrium output level. Thus in panel (ii) of Figure 3-6-1, the imputed rental income to the holders of the quota is given by area $KMN\tilde{P}$. To the extent that quota is allocated to agents outside the industry, the profitability of the industry is reduced. A CT regulator will therefore ensure that all available quota is allocated to member firms of the regulated industry. In the case of regulation by output quota then, if (3-6-1) is positive, industry profitability in panel (ii) of Figure 3-6-1 exceeds that in panel (i).

Following the analysis of Sections 2-4 and 2-5, output is declared so that the marginal expected penalty with respect to output, given the amount of output that is declared, evaluated at the regulated equilibrium output level, is equal to the tax rate. In panel (ii) of Figure 3-6-1, therefore, the unit rate of tax is equal to FB. If the industry declares R' units of output, the area $KMN\tilde{P}$ represents the size of tax payments that it incurs. From (3-4-11) the amount $(1-a)KMN\tilde{P}$ is an additional loss of potential profit within the regulated environment given that regulation is by enforced sales tax with partially declared output. Whether or not industry profitability in this case is larger or smaller than that in panel (i) of Figure 3-6-1 is not immediately apparent. It is clear however, given that $a < 1$ so that not all tax payments are refunded, that at any regulated equilibrium, if it is optimal for the CT regulator to have a non-zero quota, industry profitability with a positive and binding output quota exceeds that with a sales tax whatever the declaration strategy.[9] A CT regulator will therefore choose to regulate by output quota rather than sales tax.

Algebraically, industry profits in the situation illustrated in panel (i) of Figure 3-6-1 are

$$(3\text{-}6\text{-}2) \quad \pi_t' = \int_0^{\tilde{Q}} [h(Q)-MPC(Q)]dQ - (1-a)EP(\tilde{Q}|X=0) - b\omega(L_e')$$

and

$$(3\text{-}6\text{-}3) \quad \pi_R' = \int_0^{\tilde{Q}} [h(Q)-MPC(Q)]dQ - (1-a)EP(\tilde{Q}|R=0) - b\omega(L_e')$$

where $\pi_t'$ and $\pi_R'$ are profits in the cases of regulation by sales tax with zero output declaration and regulation by zero output quota respectively.

Examining (3-6-2) and (3-6-3), the integral term represents industry profit generated at output level $\tilde{Q}$ in the absence of any costs of the regulatory process itself. The term $b\omega(L_e')$ shows the resource cost to the industry of the enforcement level $L_e'$ necessary to generate a regulated equilibrium at $\tilde{Q}$. These two terms are common to both equations. A comparison of industry profitability within the two regulatory regimes, therefore, requires an evaluation of the relative magnitudes of the terms $EP(\tilde{Q}|X=0)$ and $EP(\tilde{Q}|R=0)$ which represent fine payments incurred under the sales tax and zero quota respectively. Given the assumption that the form of the fine function is the same in each case, as is the extent of illegal output, equations (3-6-2) and (3-6-3) are identical in value. Thus, ceteris paribus, there is no difference in industry profitability between regulation by zero quota and regulation by a sales tax set at such a level that it precludes any non-zero output declaration.

In panel (ii), industry profits are

$$(3\text{-}6\text{-}4) \quad \pi''_t = \int_0^{\tilde{Q}} [h(Q)-MPC(Q)]dQ - (1-a)EP(\tilde{Q}|X=R') - (1-a)\tau(R')-b\omega(L_e'')$$

and

$$(3\text{-}6\text{-}5) \quad \pi''_R = \int_0^{\tilde{Q}} [h(Q)-MPC(Q)]dQ - (1-a)EP(\tilde{Q}|R=R') - b\omega(L_e'')-H_Q(.)[R'-R_I]$$

where $\pi''_t$ represents profits in the case of regulation by sales tax with output declaration of R' units and $\pi''_R$ denotes profits in the case of regulation by output quota of R' units.

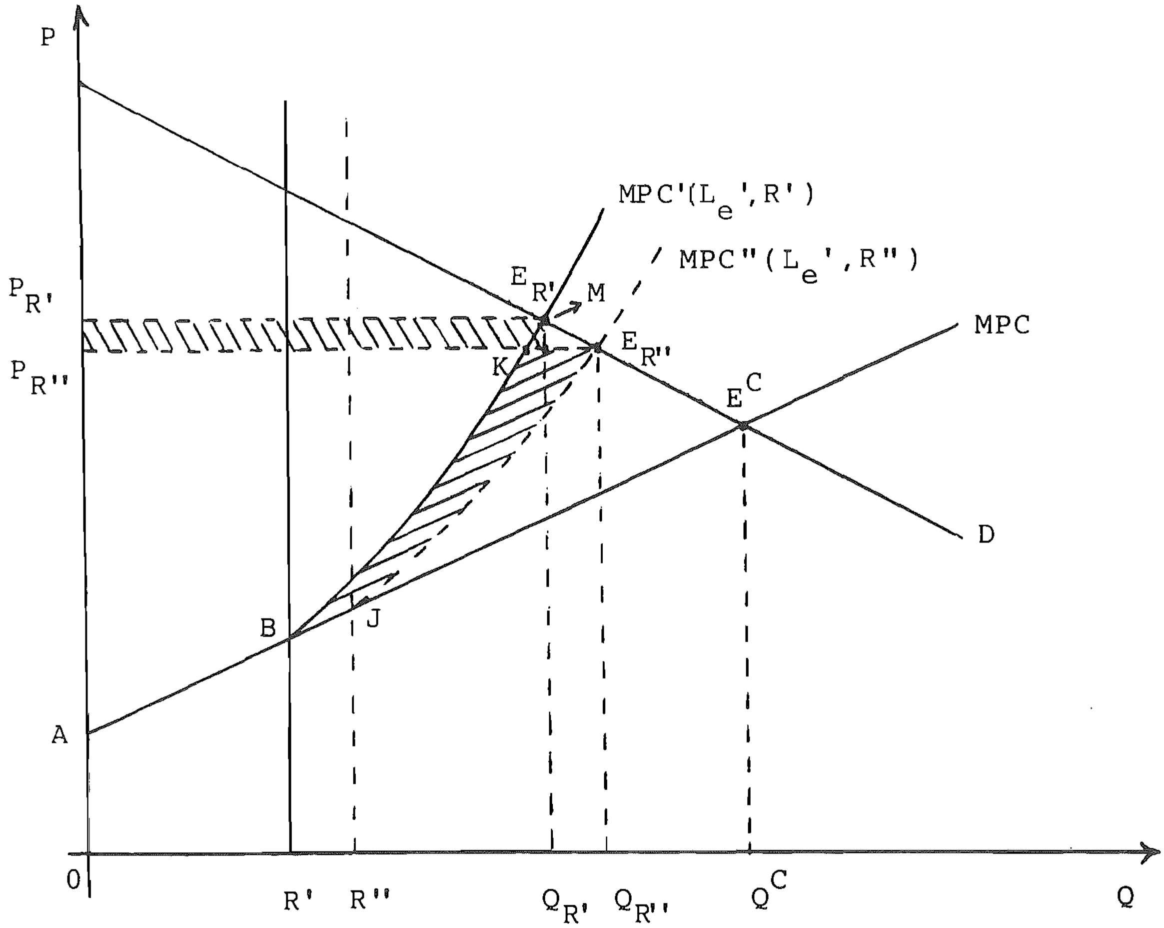
The term $(1-a)\tau(R')$ in (3-6-4) shows net tax payments incurred by the industry on the R' units of output declared. Comparing (3-6-2) and (3-6-4), the integral term is common to both equations. Following the earlier discussion the resource cost of enforcement required to generate a regulated equilibrium at $\tilde{Q}$ in panel (ii) exceeds that in panel (i). Expected fine payments, however, are unambiguously reduced by the positive output declaration. Relative profitability in the two situations therefore is not unambiguously determined.

Similarly, a comparison of (3-6-3) and (3-6-5) reveals no unambiguous relationship. Following the analysis of Section 2-6, $H_Q(.)$ is the marginal expected penalty with respect to output in the case of regulation by an output quota and reflects the imputed rental value of a unit of quota. The term $H_Q(.)[R'-R_I]$, where $R_I$ is defined as the amount of quota initially allocated to members of the industry, shows the amount of profit lost to the industry from any units of quota not initially allocated within the industry. As has been previously observed, a CT regulator will ensure that all available quota is allocated to members of the industry in order that this potential loss of profits does not occur.

Finally, comparing (3-6-4) and (3-6-5), assuming that $R_I=R'$, reveals that industry profit is unambiguously greater under regulation by non-zero output quota than under regulation by sales tax with partial output declaration which generates the same regulated equilibrium, given that the form of the expected penalty function in both cases is identical. If, therefore, a zero output quota is not optimal for a captured regulator, CT regulation, should it occur at all, will involve the use of a non-zero output quota.

Figure 3-6-1 shows changes in the level of output quota and enforcement activity such that the regulated equilibrium output level of the industry remains unaltered. Following Proposition 2-6-2, changes in available quota at any given level of enforcement which affect the marginal expected penalty with respect to output at any given illegal output level will alter the regulated industry equilibrium. This is illustrated in Figure 3-6-2 below where it is assumed that $\partial Q/\partial R > 0$.

Figure 3-6-2: The effect of a change in available quota at a given level of enforcement




With quota level R' and enforcement level $L_e'$ the industry's penalty-inclusive marginal cost curve is $MPC'(L_e',R')$ and regulated industry equilibrium occurs at $E_{R'}$ with total profits of $ABE_{R'}P_{R'}-b\omega(L_e')$. An increase in available quota to R" decreases the marginal expected penalty with respect to output at any illegal output level. The penalty-inclusive marginal cost curve, given the same amount of enforcement activity, falls to $MPC''(L_e',R'')$ and hence the regulated industry equilibrium shifts to $E_{R''}$. Regulated equilibrium output expands from $Q_{R'}$ to $Q_{R''}$. Industry profits in this situation are given by $AJE_{R''}P_{R''}-b\omega(L_e')$.

Whether or not industry profits are increased by this change in quota depends on the relative sizes of the shaded areas $P_{R''}ME_{R'}P_{R'}$ and $BJE_{R''}K$ which respectively show the decrease and increase in profits as a result of the change in quota. The change in quota size and in the marginal expected penalty with respect to output evaluated at the regulated equilibrium output level affect the value of the imputed quota rents, however, given the result that all quota is initially allocated to members of the industry, this does not affect aggregate industry profitability.

The increase in quota illustrated in Figure 3-6-2 will be instituted by a CT regulator only if it improves industry profitability. Optimal levels of quota and enforcement are determined according to the results of Propositions 3-4-1 and 3-5-1 to 3-5-4. It may be, however, for reasons that are not explicitly dealt with here, that the CT regulator is constrained to use a non-optimal instrument such as the sales tax. In this case, "optimal" regulatory policy proceeds according to the results of Propositions 3-4-2 and 3-5-5 to 3-5-8. The analysis of optimal policy contained in these Propositions is, in effect, a second-best approach for the CT regulator.

These results are now summarized in the following Proposition.

<u>PROPOSITION 3-6-1</u>: <u>Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota</u>

<u>or a strictly positive unit rate sales tax each of which is enforced by means of an expected penalty that is strictly convex in the extent of illegal output and contains no flat-rate component. Suppose also that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation and that the marginal cost of enforcement is non-decreasing. A captured regulator will choose to regulate the industry by means of a strictly binding output quota. Typically this quota will be set at a non-zero level which is strictly less than the regulated equilibrium output level of the industry. A captured regulator ensures that all available quota is allocated to members of the industry. Following Proposition 3-4-2, if enforcement is by means of a flat rate per unit expected penalty, a captured regulator will choose to regulate the industry by means of an output quota allocated entirely to members of the industry and set at the regulated equilibrium output level consistent with the CT-optimal level of enforcement $\tilde{L}_e$.</u>

## 3-7 A COMPARISON OF REGULATION UNDER NPIT AND CT HYPOTHESES

The analysis in Sections 3-4, 3-5 and 3-6 derived the characteristics of optimal regulatory policy under both the NPIT and CT hypotheses of regulatory objectives, at given values of various parameters, and examined the response of optimal policy in each case to changes in these parameters.

Under the NPIT hypothesis the regulator acts so as to maximize aggregate social welfare within the regulated environment. The CT regulator however seeks to maximize the profits of the industry within the constraints imposed by the necessity of enforcement. The CT-optimal regulated equilibrium is equivalent to that of a self-regulating cartel where formation and operating costs are partially funded by external subscription. The CT-optimal enforcement level generates profits within the regulated environment by deterring output violations at the regulated equilibrium output level. With no potential entrants to the industry, the marginal expected penalty at the regulated equilibrium output level acts to prevent chiselling by member firms of the industry. Regulation within the CT framework therefore is a solution to the internal policing problem of a cartel. To the extent that the regulator is able to garnish funds from other sectors of the economy, the internal funding requirements of forming and maintaining the cartel are reduced and hence, a regulated cartel may profitably exist where an industry-funded operation would be infeasible.

These regulatory objectives differ markedly implying that the behavioural implications for regulatory policy which result from each hypothesis may be expected to vary in several respects. Such differences may arise in the optimal choices of enforcement levels, tax rates and quota sizes, and also in the responsiveness of these choices to changes in parameter values.

In general, comparisons can not be made because of the simultaneity requirement on the relevant first-order conditions. In some cases, however, valid conclusions may be drawn on the basis of one of the relevant first-order conditions only. This depends on the cross-partial derivative term in the matrix of second-order coefficients. Several possibilities arise.

Following Proposition 3-2-1, assuming that the expected penalty is strictly convex in the extent of the violation and that control over the industry is implemented, NPIT-optimal regulatory policy involves some strictly positive level of enforcement ($L_e^*$) coupled with either a zero output quota or a sales tax, the rate of which is no less than the difference between the demand price and marginal cost of production at the regulated equilibrium consistent with the optimal level of enforcement $L_e^*$. In the case of a captured regulator, following Propositions 3-4-1 and 3-4-2, any strictly positive level of enforcement $\tilde{L}_e$ that is optimal for the CT regulator is coupled with either a sales tax ($\tilde{t}$), the rate of which does not exceed the NPIT-optimal rate $t^*$, or an output quota which is strictly less than the regulated equilibrium associated with $\tilde{L}_e$ and may be zero.

Given the assumption that the expected penalty function is strictly convex in the extent of violation, if regulation is by means of an output quota then, following Appendix 3-3, the cross derivative term $\partial^2\pi_{R_e}(.)/\partial L_e \partial R$, denoted by B in (3-5-1), is negative. That is, an increase in available quota makes the change in profits as a result of a change in enforcement smaller in absolute terms. Given the result from the above-mentioned Propositions that the optimal quota level ($\tilde{R}$) under CT is at least as great as that under NPIT ($R^*$), a result which set $\tilde{L}_e < L_e^*$ based on the consideration of the first-order condition with respect to enforcement only would occur despite the relative size of the optimal quotas rather than because of it. In fact, exclusion of the cross-derivative effect would tend to bias the conclusion towards $\tilde{L}_e > L_e^*$.

Alternatively, if regulation is by means of a unit rate sales tax then, following Appendix 3-4, the cross derivative term $\partial^2\pi_{R_e}(.)/\partial L_e \partial t$, represented by B in (3-5-24), is positive. That is, an increase in the tax rate increases in absolute terms the change in profit which occurs in response to the change in the level of enforcement. Given the result which follows from the above-mentioned Propositions that the optimal tax rate ($\tilde{t}$) under CT regulation does not exceed that under NPIT ($t^*$), a result which set $\tilde{L}_e > L_e^*$ based on the consideration of the first-order condition with respect to enforcement would occur despite the relative size of the optimal tax rate rather than because of it. Here, therefore, the bias would be towards $\tilde{L}_e < L_e^*$.

These examples represent instances whereby the conclusions of the Propositions which follow in this Section may be invalidated. There are two circumstances, however, where comparisons on the basis of the first-order condition with respect to enforcement only do provide unambiguous results. The first also occurs under the assumption of an expected penalty function that is strictly convex in the extent of the violation. If, in this case, the NPIT-optimal and CT-optimal levels of the regulatory instruments are identical, the relative size of the regulatory instruments under each policy is not relevant to the result concerning the relative sizes of optimal enforcement.

Secondly, the existence of a flat-rate per unit expected penalty allows for valid comparisons to be drawn on the basis of consideration of the enforcement first-order conditions only. This is because, as shown in Propositions 3-2-1 and 3-4-1, the NPIT-optimal quota level in this instance is boundedly indeterminate as is the tax rate while, under CT, the quota is set exactly at the regulated equilibrium output level consistent with $\tilde{L}_e$ and regulation by sales tax is not undertaken. Given the assumption that, in this case, the quota level can be instantaneously and costlessly adjusted, comparisons can again be made on the basis of the first-order condition for enforcement only.

These arguments are summarized in the following Lemma.

LEMMA-3-7-1: Given the simultaneity requirement on the relevant first-order conditions, comparisons of optimal regulatory policy under NPIT and CT hypotheses made on the basis of one of the relevant first-order conditions only are not valid in general. In particular, assuming that the regulatory instrument is enforced by means of an expected penalty that is

<u>strictly convex in the extent of the violation, the cross derivative term in the matrix of second-order conditions implies that CT-optimal enforcement levels, when comparisons are made in the absence of consideration of the other relevant first-order condition, are biased upwards relative to the NPIT-optimal level in the case of regulation by output quota and biased downwards in the case of regulation by sales tax. If, however, optimal levels of the regulatory instrument under NPIT and CT hypothesis are identical or the regulatory instrument is enforced by a flat rate per unit expected penalty, conclusions reached on the basis of one relevant first-order condition only are valid</u>.

The first potential difference to be considered concerns the optimal amount of enforcement under the two regulatory regimes. This involves a comparison of the solutions to the first-order conditions (3-2-10) and (3-4-17) or (3-4-23), where (3-2-10) shows the marginal welfare effect of enforcement in the case of NPIT regulation and (3-4-17), (3-4-23) show the marginal profitability of enforcement to the industry under CT regulation by output quota and sales tax respectively. Unfortunately, as the marginal benefits of enforcement in each situation cannot be unambiguously compared, a general condition is not forthcoming.

LEMMA 3-7-2: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax enforced by means of an expected monetary penalty that is convex in output and the extent of illegal behaviour. Given that the marginal resource cost of an extra unit of enforcement is strictly positive and exceeds any marginal revenues that it generates through the payment of fines and/or taxes, the following comparisons can be made:

(i) If the proportion of the resource cost of enforcement that the industry is required to fund is no greater than the proportion of fine payments, and when applicable tax revenues, that are refunded to members of the industry, then the marginal cost of a given level of enforcement is unambiguously greater for the NPIT regulator than the CT regulator.

(ii) If, however, the proportion of resource costs that are funded by the industry exceeds the proportion of fine payments, and where applicable tax revenues, that are refunded to the industry, then it

<u>is possible that the marginal cost of a given level of enforcement is greater for the CT regulator than for the NPIT regulator</u>.

<u>Proof</u>.

(i) In the case of NPIT regulation the marginal cost of enforcement is given solely by the marginal resource cost $\omega'(L_e)$. In the case of CT regulation, however, the marginal cost of enforcement consists of the marginal fine payments, and where applicable marginal tax payments, that are incurred by the industry in addition to the proportion of marginal resource cost that it is required to fund.

Following (3-4-7) and (3-4-12) a composite function $EP_\Omega(L_e,\Omega,\mu)$ is taken to represent the additional costs incurred by the industry in the CT regulatory environment. This follows from the terminology of Sections 3-2 and 3-3 where $\Omega$ is a proxy for the regulatory instrument which can be either a quota or sales tax. In the case of regulation by sales tax, the expression includes tax payments and in both cases, $\mu$ represents the state of enforcement technology. Here, the special case, when funding of enforcement activity and disbursement of revenue proceeds on a per capita proportional basis, is used. It can be shown that this causes no loss of generality in the result but merely simplifies the algebra.[10]

Assume that the opposite conclusion to that presented in the Lemma holds. Then, using (3-2-2) and

(3-4-13), and the composite function ($EP_\Omega(L_e, \Omega, \mu)$

$$(3\text{-}7\text{-}1) \qquad \frac{z}{m}\,\omega'(L_e) + [1-\frac{z}{m}]\frac{\partial EP_\Omega}{\partial L_e} > \omega'(L_e)\,; \qquad \frac{\partial EP_\Omega}{\partial L_e} > 0,\ \omega'(L_e) > 0$$

Subtracting $\partial EP_\Omega(.)/\partial L_e$ from both sides of (3-7-1) and simplifying gives

$$(3\text{-}7\text{-}2) \qquad \frac{z}{m}[\omega'(L_e) - \frac{\partial EP_\Omega}{\partial L_e}] > \omega'(L_e) - \frac{\partial EP_\Omega}{\partial L_e}$$

where both sides of (3-7-2) are positive by the assumption that the marginal resource cost of enforcement exceeds any additional revenues raised. The inequality in (3-7-2) therefore holds if and only if $z > m$. This however contradicts (3-4-13) and hence the conclusion in the Lemma must hold.

(ii) From Section 3-4, 'b' represents the proportion of the resource cost of enforcement that is funded by the industry and 'a' the proportion of fines and/or taxes that are refunded to members of the industry. The marginal cost of enforcement to the CT regulator denoted here by $w_\Omega$ is then

$$(3\text{-}7\text{-}3) \qquad w_\Omega = b\omega(L_e) + (1-a)EP_\Omega$$

Using (3-7-3) and (3-2-2), $\partial w_\Omega/\partial L_e \gtreqless \partial\omega(L_e)/\partial L_e$ if and only if

$$(3\text{-}7\text{-}4) \qquad b\omega'(L_e) + (1-a)\frac{\partial EP_\Omega}{L_e} \gtreqless \omega'(L_e)$$

and rearranging, $\partial w_\Omega/\partial L_e \gtreqless \partial\omega(L_e)/\partial L_e$ if and only if

$$(3\text{-}7\text{-}5) \qquad \omega'(L_e) \lesseqgtr [\frac{1-a}{1-b}]\frac{\partial EP_\Omega}{\partial L_e}$$

Given the assumption that $b > a$, then $[(1-a)/(1-b)] > 1$ and the marginal cost of enforcement will be greater to the CT regulator than to the NPIT regulator unless the marginal resource cost is sufficiently larger than the additional payments incurred.

□

For the purposes of the following Proposition the social cost function from (3-2-8) is modified to include an additional shift parameter. Thus

$$(3\text{-}7\text{-}6) \qquad SC = SC(Q(L_e, \Omega, \mu, Y), \phi, \psi) \ ; \ SC_3 > 0, \ SC_{13} > 0$$

where $\psi$ is the additional shift parameter. An increase in $\psi$ increases marginal social cost at any output level.

PROPOSITION 3-7-1: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax each of which is enforced by means of an expected monetary penalty that is convex in output and the extent of illegal behaviour. Given the caveats expressed in Lemma 3-7-1:

(i) If enforcement is costless to both the NPIT and CT regulators then the NPIT-optimal level of enforcement ($L_e^*$) is greater than, less

<u>than, or equal to the CT-optimal level of enforcement ($\tilde{L}_e$) if and only if the first-best socially optimal output level ($\hat{Q}$) is less than, greater than, or equal to the monopoly profit maximizing output level ($Q^M$).</u>

(ii) <u>If the marginal cost of enforcement to the NPIT regulator exceeds that to the CT regulator and both are strictly positive such that it is optimal for enforcement to occur under each hypothesis, then there is some degree of negative externality such that the NPIT-optimal level of enforcement ($L_e^*$) is greater than, less than, or equal to the CT-optimal enforcement level ($\tilde{L}_e$) if and only if the negative externality generated by the industry is more severe, less severe, or identical to this critical value.</u>

(iii) <u>If the marginal cost of enforcement to the NPIT regulator is less than to the CT regulator then the degree of externality required before NPIT-optimal enforcement exceeds CT-optimal enforcement is smaller than if the marginal cost</u>

<u>of enforcement to the NPIT regulator exceeds that to the CT regulator. In particular, it is possible in these circumstances that the presence of any negative externality results in greater NPIT-optimal enforcement than CT-optimal enforcement</u>.

<u>Proof</u>.

(i) Given the assumption that the unregulated competitive equilibrium output level is socially excessive and exceeds the profit-maximizing monopoly output $Q^M$, enforcement occurs under each hypothesis only if $\partial Q/\partial L_e < 0$. This follows from previous analysis given the assumptions that the expected penalty function is convex in the extent of illegal behaviour and that additional enforcement activity increases the marginal expected penalty with respect to output at any illegal output level.

From (3-2-10) the NPIT-optimal level of enforcement, in a world of costless enforcement, is that level $L_e^*$ which generates a regulated equilibrium at the first-best socially optimal output level $\hat{Q}$ such that, using the modified social cost function from (3-7-6)

$$h(\hat{Q}(L_e^*, \Omega, \mu), Y) - SC_1(\hat{Q}(L_e^*, \Omega, \mu), \phi, \psi) = 0 \tag{3-7-7}$$

where $\psi$ is an additional shift parameter incorporated in the social cost function. An increase in $\psi$ increases marginal social cost at any output level.

Substituting $L_e^*$ into (3-4-17) or (3-4-23) using the composite formulation, given that the marginal cost to the industry of enforcement is strictly positive, gives

$$(3\text{-}7\text{-}8)\quad \left.\frac{\partial \pi_{R_e}}{\partial L_e}\right|_{L_e=L_e^*} = [h_1(Q(L_e^*,\Omega,\mu),Y)Q(L_e^*,\Omega,\mu)$$

$$+ h(Q(L_e^*,\Omega,\mu),Y) - C'(Q(L_e^*,\Omega,\mu))]\frac{\partial Q}{\partial L_e}$$

Given that $\partial Q/\partial L_e < 0$, $\left.\frac{\partial \pi_{R_e}}{\partial L_e}\right|_{L_e=L_e^*} \gtrless 0$ if and only if

$$(3\text{-}7\text{-}9)\quad [h_1(Q(L_e^*,\Omega,\mu),Y)Q(L_e^*,\Omega,\mu) + h(Q(L_e^*,\Omega,\mu),Y)$$

$$- C'(Q(L_e^*,\Omega,\mu))] \lessgtr 0$$

The CT-optimal enforcement $\tilde{L}_e$ occurs where $\partial\pi_{R_e}/\partial L_e = 0$ which, given that $\partial Q/\partial L_e < 0$, requires that the bracketed term of (3-7-9) evaluated at $\tilde{L}_e$ be equal to zero. This corresponds to the monopoly profit-maximizing output level where marginal revenue equals marginal cost.

Taking the case where the bracketed term in (3-7-9) is negative, (3-7-9) shows that it is optimal for the CT regulator to expand the enforcement effort. This implies that $\tilde{L}_e > L_e^*$. Given that, as illustrated in Figure 3-4-1, the absolute value of the bracketed term is increasing in output over the relevant output range, it follows in this case that the CT-optimal regulated equilibrium output level $Q^M$ occurs at a lower output level than $\hat{Q}$ and hence the result holds.

The results for the other cases can be established using a similar argument.

(ii) Taking $L_e^* > 0$ as the NPIT-optimal level of enforcement such that, from (3-2-10) $\partial W_{R_e}/\partial L_e^* = 0$, and substituting this into the composite form of the CT regulator's first-order condition with respect to enforcement using the special case of per capita proportional funding and redistribution, which from Lemma 3-7-2 part (i) creates no loss of generality, $\tilde{L}_e \gtreqless L_e^*$ if and only if

$$(3\text{-}7\text{-}10) \quad [h_1(Q(L_e^*,\Omega,\mu),Y)Q(L_e^*,\Omega,\mu) + h(Q(L_e^*,\Omega,\mu),Y) - C'(Q(L_e^*,\Omega,\mu))]\frac{\partial Q}{\partial L_e} - \left[\frac{z}{m}\omega'(L_e)+[1-\frac{z}{m}]\frac{\partial EP_\Omega(L_e^*,\Omega,\mu)}{\partial L_e}\right] - \left[[h(Q(L_e^*,\Omega,\mu),Y)-SC_1(Q(L_e^*,\Omega,\mu),\phi,\psi)]\frac{\partial Q}{\partial L_e} - \omega'(L_e)\right] \gtreqless 0$$

Rearranging and simplifying (3-7-10) reveals the condition that $\tilde{L}_e \gtreqless L_e^*$ if and only if

$$(3\text{-}7\text{-}11) \quad SC_1(Q(L_e^*,\Omega,\mu),\phi,\psi)-C'(Q(L_e^*,\Omega,\mu)) \lesseqgtr \left[\frac{z}{m}\omega'(L_e) + (1-\frac{z}{m})\frac{\partial EP_\Omega(L_e^*,\Omega,\mu)}{\partial L_e} - \omega'(L_e)\right]\frac{\partial Q}{\partial L_e} - h_1(Q(L_e^*,\Omega,\mu),Y)Q(L_e^*,\Omega,\mu)$$

From part (i) of Lemma 3-7-2 the bracketed term on the right-hand side of (3-7-11) is negative as, from (A3-1-1) and (3-2-10), are $h_1(Q(L_e^*,\Omega,\mu), Y)$ and $\partial Q/\partial L_e$ respectively. Given these conditions, the right-hand side is some positive number.

Assume that $\overline{\psi}$ is the value of the shift parameter such that its associated marginal social cost level $SC_1(L_e^*,\Omega,\mu),\phi,\psi)$ is that which gives equality in (3-7-11). From (3-7-11) then, given $\psi = \overline{\psi}$, $\tilde{L}_e$ is also the NPIT-optimal level of enforcement and thus $L_e^*\big|_{\overline{\psi}} = \tilde{L}_e$.

An increase in $\psi$, following the assumption in 3-7-6, results in a higher marginal social cost at any given output level. This enlarges the discrepancy between marginal social cost and marginal private cost of output produced in the industry and corresponds to an increase in the severity of the negative externality generated by the industry. Any increase in $\psi$ then, to some $\overline{\psi} > \overline{\overline{\psi}}$, increases the value of the left-hand side of (3-7-11) so that it exceeds the right-hand side as the right-hand side is unchanged. Expression (3-7-11) then implies that the NPIT regulator can raise regulated aggregate surplus by increasing the amount of enforcement activity and thus $L_e^*\big|_{\overline{\overline{\psi}}} > \tilde{L}_e$. The converse result can be established by a similar argument with the opposite assumption on the size of $\psi$.

(iii) If the marginal cost of enforcement to the NPIT regulator is less than that to the CT regulator, the bracketed term on the right-hand side of (3-7-11) is positive. Examining the right-hand side of (3-7-11) reveals that it is greater than, less than, or equal to zero if and only if

$$(3\text{-}7\text{-}12)\quad \left[\frac{z}{m}\ \omega'(L_e)+[1-\frac{z}{m}]\frac{\partial EP_\Omega(L_e^*,\Omega,\mu)}{\partial L_e}-\ \omega'(L_e)\right]\frac{\partial Q}{\partial L_e}$$

$$\gtreqless h_1(Q(L_e^*,\Omega,\mu),Y)Q(L_e^*,\Omega,\mu)$$

Following the assumptions and conditions on the signs of the component terms of (3-7-11) both sides of (3-7-12) are negative. Assuming that the 'greater than' inequality holds in (3-7-12), the right-hand side of (3-7-11) is some positive number. Given that the bracketed term on the right-hand side of (3-7-11) is positive, however, the magnitude of the right-hand side of (3-7-11) here is less than that in part (ii) of the Proposition where it was assumed that the bracketed term was negative. Using this and the assumption in 3-7-6, the critical value of $\psi$ that generates equality in (3-7-11) occurs at some $\bar{\bar{\bar{\psi}}} < \bar{\psi}$. Thus the degree of externality required before NPIT-optimal enforcement exceeds the CT-optimal level, if the marginal cost of enforcement to the CT regulator exceeds that to the NPIT regulator, is in general less than if this inequality is reversed.

If the 'less than' inequality holds in (3-7-12) then the right-hand side of (3-7-11) is negative. The presence of any negative externality implies that marginal social cost exceeds marginal private production cost and in these circumstances ensures that, following (3-7-11), the NPIT-optimal amount of enforcement is unambiguously greater than that which is optimal for a CT regulator.

□

Proposition 3-7-1 shows that, as may have been expected, the CT-optimal level of enforcement is in general different from that which is optimal for the NPIT regulator. Result (i) of the Proposition is intuitively obvious. In a world of costless enforcement, it is optimal to expand enforcement to the point where the marginal benefit of so doing is zero. As previously defined the resulting output levels are $\hat{Q}$ and $Q^M$ for the NPIT and CT regulators respectively and the result follows from equation (3-2-5) given that $\partial Q/\partial L_e < 0$.

Enforcement costs are introduced in parts (ii) and (iii) of the Proposition using the results of Lemma 3-7-2. In (3-7-11) each component term is fixed and given by the demand and cost conditions in the market except for the marginal social cost of output. The condition expressed in (3-7-11) implies that the more severe is the negative externality, the greater will be the NPIT-optimal control of the industry in comparison with that under CT regulation. This is in effect a generalization of result (i) of Proposition 3-7-1 allowing for the existence of some strictly positive cost structure for enforcement activity. An increase in the parameter $\psi$, which corresponds to a higher level of marginal social cost at any given output level and hence reflects an increase in the severity of the negative externality, has the effect of decreasing the first-best socially optimal output level $\hat{Q}$ relative to a fixed $Q^M$ and allows for an increase in the size of NPIT-optimal enforcement.

One illustrative polar case is where there is no externality so that marginal social cost is identical to the marginal private cost of production. Here the competitive equilibrium is socially optimal and thus NPIT-optimal enforcement is zero but the CT regulator may well seek to restrict industry output towards the monopoly profit-maximizing level with some positive amount of enforcement activity. Beginning from this situation then, the emergence of a negative externality allows for the possibility of strictly positive enforcement activity by the NPIT regulator and the greater is the degree of externality, the higher is $L_e^*$ relative to the fixed amount of CT-optimal enforcement.

Unfortunately, from the perspective of inferring regulatory objectives from observed behaviour, the quantitative differences in enforcement activity which emerge from Proposition 3-7-1 are not unique and more significantly, except for the particular circumstances outlined in part (iii) of the Proposition, or where there is no externality, the qualitative nature of the differences are not unambiguous. To know whether or not the level of enforcement activity reveals a captured regulator requires enough knowledge to solve the optimization problems of the NPIT and CT regulators.

The second behavioural difference between NPIT and CT regulation to be considered here occurs in the choice of regulatory instrument. This chapter has examined the effects of regulation by output quota and unit rate sales tax, each of which is enforced by an expected monetary penalty. The analysis of Sections 2-4, 2-5 and 2-6 showed that it is possible to generate the same regulated equilibrium output using either instrument. The distributional impacts of the two instruments are markedly different.

A NPIT regulator is assumed not to be interested in the distributional impacts of regulatory policy. Any fine revenues, tax payments or quota rentals constitute transfer payments within the economy and as such are irrelevant to the decisions of the NPIT regulator. Following Proposition 3-2-1, if the expected penalty function is strictly convex in the extent of illegal behaviour, no strictly positive quota level is optimal for the NPIT regulator. This is because any quota greater than zero increases the resource cost of achieving a given regulated equilibrium output level. The zero quota then minimizes the resource cost of regulation and the expected penalty is equivalent to a non avoidable tax on output.

A regulated industry equilibrium occurs where the marginal expected penalty with respect to output is equal to the difference between the demand price and marginal private cost of production. If the expected penalty is a constant flat-rate amount per unit of output then the marginal expected penalty with respect to output, and hence the regulated equilibrium output level of the industry, is unaffected by any change in the level of available quota provided that the quota remains binding on behaviour. With this form of expected penalty function therefore, the NPIT regulator is indifferent between any levels of output quota that do not exceed the regulated equilibrium output level of the industry. If a non-zero quota is used by the NPIT regulator, in these circumstances, there is no clear pattern of quota allocation that would be expected to emerge.

Given that the quota is a valuable asset, if the NPIT regulator is concerned with the distributional impacts of its regulatory policies, not all of the available quota will be allocated to members of the regulated industry.

The CT regulator, in contrast to the NPIT regulator, is particularly concerned with the distributional impacts of its policies as they affect the profits of the regulated industry. Proposition 3-6-1 states that a CT regulator will choose to regulate the industry by output quota. There the argument is established using an expected penalty that is strictly convex in the extent of illegal output and contains no flat-rate component. The same argument can also be used to establish the result if the expected penalty is a constant flat-rate amount per unit of output. In this case there is a one to one correspondence between the change in quota rentals and the change in expected penalty payments. Given that the expected penalty is some constant flat-rate amount per unit, a change in the quota level, provided that it remains binding on behaviour, will not alter the marginal expected penalty with respect to output and hence, from Proposition 2-7-2, does not affect the regulated equilibrium output level of the industry. In these circumstances then, as shown in Figure 2-7-1, a unit change in available quota serves to directly transfer the value of the marginal expected penalty between quota rents and fine payments. As shown in Proposition 3-4-1, it is therefore optimal for the CT regulator to set the quota at the regulated equilibrium output level that is consistent with the optimal enforcement level $\tilde{L}_e$ and, following Proposition 3-6-1, all of this quota is allocated to members of the industry.

This example of a constant unit rate marginal expected penalty illustrates that the existance of a non-zero binding output quota allocated entirely to members of the regulated industry prevents leakage of potential profits from the industry by enabling member firms to, in effect, pay fines directly to themselves thus avoiding the distribution of these payments throughout the economy. Under a sales tax regime, fine payments can be avoided only at the expense of incurring taxes which also constitute a loss of potential profits. The choice of output quota over sales tax as the optimal regulatory instrument by the CT regulator is another example of the thesis that producers prefer direct controls to taxes [Buchanan and Tullock, 1975].

These results are summarized in the following Proposition:

PROPOSITION 3-7-2: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax each of which is enforced by an expected monetary penalty. Suppose also that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation and that the marginal cost of enforcement is non-decreasing.

<u>If the expected monetary penalty is strictly convex in the extent of illegal output and contains no flat rate component, the NPIT regulator is indifferent between a zero quota and a tax rate set at a level such that only the marginal unit of output, if any, is declared. The NPIT regulator will not employ a non-zero output quota. If, however, the expected penalty is some constant flat-rate amount per unit of illegal output, the NPIT regulator is indifferent between a tax rate set at a level not less than the marginal expected penalty with respect to output and any level of output quota which remains binding on behaviour. In this case no unique pattern of quota allocation would be expected to emerge.</u>

<u>If the expected monetary penalty is strictly convex in the extent of illegal output and contains no flat-rate component, a CT regulator is also indifferent between a zero quota and a tax rate set at a level such that only the marginal unit of output, if any, is declared. If enforcement costs to the industry are such that it is profitable to do so, a CT regulator will employ a non-zero binding output quota in preference to</u>

either of these strategies and also in preference to any other tax rate consistent with the same regulated equilibrium output level as the quota, whatever the optimal declaration response to this tax may be. If, however, the expected monetary penalty is some constant flat-rate amount per unit of output, the CT regulator will set the quota at the regulated equilibrium output level of the industry consistent with the CT-optimal enforcement level so that no violations of the constraint will occur and hence no fines are incurred by the industry. In each case, when a non-zero quota is employed, the CT regulator ensures that the available quota is allocated entirely to member firms of the industry. The existence of a non-zero quota, given that the expected monetary penalty is strictly convex in the extent of illegal output and contains no flat-rate component, reveals unambiguously that the regulatory agency is captured. If, however, the expected monetary penalty is some constant flat-rate amount per unit of illegal output, the existence of an output quota which is allocated entirely to members of the regulated industry and set at a level such that no violations

<u>occur, and hence no fines are levied, is an almost certain indication that the regulatory agency is captured</u>.

The final source of behavioural differences between NPIT and CT regulators to be considered here is found in their responses to parameter changes. Two such changes which affect the optimal regulatory policy of both types of regulator were analysed in Sections 3-3 and 3-5. These were changes in the level of aggregate consumer income and changes in the state of enforcement technology. Examination of Propositions 3-3-1, 3-3-3, 3-5-1, 3-5-4, 3-5-5 and 3-5-8 reveals that a comparison of optimal responses by NPIT and CT regulators to changes in either of these parameters, as in the case of the comparison of optimal enforcement levels in Proposition 3-7-1, requires enough information to solve the optimization problems of the respective regulators. This is not particularly useful in terms of observational inference and hence these particular changes will not be analysed further. There are, however, particular parameters of which changes in value reveal clear differences in regulatory behaviour that reflect the motives of the regulator.

As has been previously assumed, the NPIT regulator, motivated by aggregate welfare, is unaffected by distributional effects whereas the CT regulator seeking to maximize industry profit is primarily concerned with the distributional aspects of regulatory policy. Although not specifically analysed therefore, NPIT regulatory behaviour is invariant to changes in distributional parameters such as the proportion of the resource cost of enforcement funded by the industry or the proportion of fine payments redistributed to members of the industry. Following Propositions 3-5-2, 3-5-3, 3-5-6, and 3-5-7, however, it is evident that the optimal degree of CT regulation is determined by the level of such parameters. These results are summarized in the following Proposition.

PROPOSITION 3-7-3: Suppose that an industry comprising a finite number of identical competitive firms with concave profit functions faces a market demand curve with non-zero finite elasticity and is subject to either a strictly binding output quota or strictly positive unit rate sales tax each of which is enforced by means of an expected monetary penalty which is strictly convex in the extent of illegal output. Suppose also that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation, and that the marginal resource cost of enforcement is non-decreasing. In general the response of regulatory behaviour to changes in parameter values will differ according to the objectives of the regulator. In particular, a change in the extent of regulatory control of the industry in response to a change in the value of a distributional parameter reveals the existence of a captured regulator.

## 3-8 MONOPOLY

As stated in Section 2-2, it is possible that a profit-maximizing monopolist, protected by barriers to entry and operating at $Q_i^M$ on Figure 2-2-1, may be producing an output level which exceeds the first-best level $\hat{Q}$ in Figure 2-2-2.

With a negatively sloped market demand curve and a strictly convex production cost function, the monopolistic optimum $Q^M$ occurs at an output level which is strictly less than that at the unregulated competitive equilibrium $Q^C$. Assuming that marginal social cost exceeds marginal private cost and that it also increases in output at a faster rate than marginal private cost, monopolistic control of the industry results in a smaller loss of aggregate social surplus than at the unregulated competitive equilibrium.

The fact that there is a loss of aggregate surplus, however, provides the impetus for NPIT regulation of the industry provided that the cost of so doing is not prohibitive. A necessary condition for this to occur, from (3-2-10) is that

$$\omega'(0) < [h(Q^M,Y)-SC_1(Q^M,\phi)]\frac{\partial Q}{\partial L_e} \tag{3-8-1}$$

where the left-hand side of (3-8-1) is the marginal resource cost of the initial unit of enforcement and the right-hand side is the marginal gain in aggregate surplus that it generates.

Provided that the inequality in (3-8-1) is satisfied, following (3-2-2) and (3-2-10), NPIT-optimal regulated industry equilibrium will occur at some intermediate output level $Q^{*M}$ where

(3-8-2) $\hat{Q} < Q^{*M} < Q^{M}$ [11]

<u>LEMMA 3-8-1:</u> <u>Suppose that an industry comprising a finite number of identical firms with concave profit functions which faces a market demand curve with non-zero finite elasticity is operating at the pure monopoly profit-maximizing output level ($Q^M$) which exceeds the first-best socially optimal output level ($\hat{Q}$). Suppose also that this industry is subject to either a strictly binding output quota or a strictly positive unit rate sales tax each of which is enforced by means of an identical expected monetary penalty function that is strictly convex in the extent of both actual and illegal output. Assuming that the unregulated monopoly output level $Q^M$ is less than the unregulated competitive equilibrium output level $Q^C$, and that an increase in enforcement increases the marginal expected penalty with respect to output at any given illegal output level, if the price responsiveness of demand does not decrease with an increase in quantity, the marginal effectiveness as a deterrent of an additional unit of labour devoted to enforcement at the pure monopoly profit-maximizing output level does not exceed that at the unregulated competitive equilibrium output level.</u>

Proof.

Under competitive operation of the industry a regulated equilibrium is determined at that output level where the marginal expected penalty with respect to output equates the difference between marginal private production cost and the demand price. With monopolistic control the regulated equilibrium occurs at the output level where the marginal expected penalty with respect to output equates the difference between marginal private production cost and marginal revenue.

Given that the market demand curve exhibits non-zero finite elasticity, the rate of change of marginal revenue exceeds that of demand price at any given quantity. Given the additional assumption that the price responsiveness of demand does not decrease with an increase in quantity, the rate of change of marginal revenue at any given quantity also exceeds that of the demand price at all greater quantities. Following this result, and the assumption that $Q^M < Q^C$, the rate of change in marginal revenue at $Q^M$ exceeds the rate of change in the demand price at $Q^C$.

The assumptions that the expected monetary penalty is strictly convex in output and the extent of the violation and that an increase in enforcement raises the marginal expected penalty with respect to output at any given illegal output, together with the assumption that $Q^M < Q^C$, imply that the extent of increase in the marginal expected penalty with respect to output at $Q^M$ is no greater than that at $Q^C$. Using, for the sake of argument, the most pessimistic case that the change in the marginal expected penalty is the same at both output levels,together with the above result concerning the relative rates of change of marginal revenue and demand price, it follows that the reduction in regulated equilibrium quantity which results from a unit increase in enforcement is smaller at $Q^M$ than at $Q^C$. This proves the Lemma.

□

Given the other assumptions of Lemma 3-8-1, the condition that the price responsiveness of demand does not decrease with an increase in quantity is sufficient but not necessary to ensure the result. If this condition does not hold, the validity of the result depends on a comparison of the relative rates of change of marginal revenue and the demand price and the relative magnitudes of the enforcement-induced increase in the marginal expected penalty with respect to output.

<u>PROPOSITION 3-8-1</u>: <u>Suppose that an industry comprising a finite number of identical firms with concave profit functions which faces a market demand curve with non-zero finite elasticity is operating at the pure monopoly profit-maximizing output level ($Q^M$) which exceeds the first-best socially optimal output level ($\hat{Q}$). Suppose also that this industry is subject to either a strictly binding output quota or a strictly positive unit rate sales tax each of which is enforced by means of an expected monetary penalty that is strictly convex in the extent of</u>

<u>illegal output, that regulating the output of the industry necessitates the use of scarce resources to enforce the regulation, and that the marginal resource cost of enforcement is non-decreasing. Assuming, in addition, that the price responsiveness of demand does not decrease with an increase in quantity and that the same form of the expected penalty function is used in each case, the NPIT-optimal level of enforcement activity under monopoly ($L_e^{*M}$) is strictly less than that ($L_e^{*C}$) in the case when the industry is operating competitively in the unregulated environment.</u>

<u>Proof</u>.

NPIT-optimal enforcement is given by the solution to (3-2-10) given the optimal value of the regulatory instrument such that (3-2-10) and (3-2-11) are satisfied simultaneously. Implicit in the Proposition is the assumption that it is optimal for the NPIT regulator to enforce the regulation when the industry is operating competitively so that $L_e^{*C} > 0$.

Following Proposition 3-2-1 the NPIT-optimal quota level is zero while the NPIT-optimal sales tax is set at a rate no less than that which equates the difference between market price and marginal private production cost at the output level consistent with the NPIT-optimal level of enforcement and can be costlessly adjusted within this bound.

Given the assumption that both $Q^C$ and $Q^M$ are socially excessive, the marginal social cost of production exceeds the demand price at each of these levels. Given also that the pure monopoly profit-maximizing output level $Q^M$ is less than $Q^C$ and the assumption in (3-2-8) that marginal social cost is non-decreasing in output, it follows that

$$(3\text{-}8\text{-}3) \qquad [h(Q^M,Y) - SC_1(Q^M,Y)] > [h(Q^C,Y) - SC_1(Q^C,Y)]$$

where both sides of (3-8-3) are negative. Given that the same form of expected penalty function is used in both cases, following Lemma 3-8-1,

$$(3\text{-}8\text{-}4) \qquad \left.\frac{\partial Q}{\partial L_e}\right|_{Q=Q^M} \geq \left.\frac{\partial Q}{\partial L_e}\right|_{Q=Q^C}$$

and for enforcement to occur in each case, both sides of (3-8-4) must also be negative. Finally, the resource cost of enforcement at any level of enforcement activity is the same irrespective of whether the industry is operated monopolistically or competitively.

From (3-2-10), NPIT-optimal enforcement occurs where the marginal benefit and marginal cost of enforcement to the regulator are equated. Given the above results, the marginal benefit of any level of enforcement activity, assuming that it is positive, is less at the unregulated monopolistic equilibrium output level of the industry than it is at the unregulated competitive equilibrium output level. Therefore, with a smaller marginal benefit and identical marginal cost of enforcement, the NPIT-optimal level of enforcement activity is less in the case when the industry is operated monopolistically, in the unregulated environment, than when it is operated competitively.

□

Following Proposition 3-8-1

(3-8-5) $L_e^{*M} < L_e^{*C}$

and the degree of NPIT-optimal enforcement will be less in the event of monopolistic control of the industry than under competition. This is because the social cost of the externality is smaller than that caused by competitive behaviour while the resource cost of any given level of enforcement is the same in each case. With the exception of this, all other behavioural implications of NPIT regulation, which follow from equations (3-2-10) and (3-2-11), hold in the case of monopoly just as for the competitive case previously considered.[12]

Monopolistic control of an industry involves self regulation in restricting output to the profit-maximizing level $Q^M$. Any additional regulation which would lower output from this point, while incurring the additional costs of enforcement funding and net expected penalty payments, reduces the monopolist's profit and will not be undertaken voluntarily by a CT regulator.

There are situations,however,where a CT regulator may well regulate the operations of a monopolist. One such instance is that of 'strategic' regulation. Here the monopolist is threatened with the prospect of hostile NPIT-motivated regulation and engages in strategic self regulation to preclude this possibility.[13]

As any reduction in output in this situation reduces the monopolist's current profits, the behavioural implications, following Proposition 2-8-7, are that any CT regulation will be by means of a quantitative restriction set at the maximum level possible with the minimum enforcement effort sufficient to ensure that NPIT regulation does not occur.

In this way the monopolist acts to minimize the expected loss of profits from regulatory activity. The greater is the quota, within the constraint that it must not exceed $Q^M$, the smaller is the penalty incurred at any level of enforcement. Minimizing enforcement activity minimizes net fines paid at any given violation and reduces the resource cost to the monopolist. Optimum CT enforcement under monopoly is zero which raises the spectre of voluntary restraints, however, positive expenditures may be necessary to create credible deterrents against the implementation of the NPIT threat. Anything other than token regulation of a monopolist in these circumstances then is an indication that regulatory control is not motivated by the objectives of an unfettered CT regulator.[14]

## 3-9 CONCLUSION

The analysis in Chapter Two was concerned with developing a model to analyse the effects of regulating a competitive, negative-externality generating industry in a partial equilibrium framework. Specific recognition was made of the necessity for enforcement and the differing characteristics of regulation by sales tax and by output quota were derived. The present chapter has extended the analysis of Chapter Two by incorporating, within this otherwise identical framework, two competing hypotheses concerning regulatory objectives; the NPIT approach assuming that the regulator acts so as to maximize aggregate welfare and the CT hypothesis, the pure form of which assumes that the regulator seeks to maximize regulated industry profits.

In Chapter Two it was discovered that various tax/expected penalty and quota/expected penalty combinations could be formulated to generate identical aggregate results but that the distributional implications of such policies differed. These distributional aspects of regulatory policy assume great significance when the objectives of the regulator are discussed and form the basis of the differential behavioural implications for regulatory policy derived in this chapter.

The analysis in this chapter began, under each hypothesis, by determining the optimal regulatory policy in terms of the type and level of regulatory instrument implemented and the level of enforcement employed. Following Proposition 3-2-1 a NPIT regulator concerned solely with aggregate social welfare will not, in circumstances where the level of the quota directly affects regulated equilibrium output, regulate the industry by means of a non-zero output quota. This result is in accordance with economists' agreement on the "superior efficacy of penalty taxes as instruments for controlling significant external diseconomies" [Buchanan and Tullock, 1973; 139].

By contrast, the analysis of Sections 3-4 and 3-6 reveal that a CT regulator favours regulation by output quota and is furthermore concerned about the level of such a quota. Indeed, Proposition 3-7-2 demonstrated that the existence of a non-zero output quota which is allocated entirely to members of the regulated industry and which, in the case of a constant flat-rate per unit marginal expected penalty, is set at the regulated equilibrium output level of the industry so that no violations occur and no penalties are imposed, reveals the existence of a captured regulator. Other behavioural differences also exist.

Firstly, following Proposition 3-7-1, the NPIT-optimal level of enforcement will, in general, differ from that which is optimal for a CT regulator. Unfortunately neither the quantitative nor the qualitative nature of this difference is unambiguous and hence to infer anything about the objectives of the regulator from an observation of the level of enforcement activity requires enough knowledge to solve the optimization problem of each regulator.

Secondly, the behavioural response in optimal regulatory policy to changes in various parameters also differs according to the objective of the regulator. The complexity of responses to changes in parameters such as the level of aggregate consumer income and the state of enforcement technology is such as to preclude their use for observational inference. Other responses however are more clear-cut. For instance, as stated in Proposition 3-7-3, a change in regulatory behaviour in response to an alteration in a distributional parameter, such as the proportion of the resource cost of enforcement that is funded by the regulated industry itself, is an indication that the regulator is captured. Together these behavioural characteristics provide potentially observable and testable empirical criteria by which the performance of a regulator can be monitored and by which its objectives may be inferred.

This chapter has analysed regulation in the case when the regulated industry comprises a finite number of competitive firms. Regulation by a CT regulator has been stated to be analagous to the formation of an industry cartel. Alternatively, with a variable number of firms or free entry, a CT regulator may use regulation as a barrier to entry and so facilitate a natural monopoly which would not exist in an unregulated environment.[15] The function of the regulatory instrument and enforcement in this instance is to artificially inflate the cost structures of potential competitors. This can be accomplished by allocating quotas to incumbent firms, or some subset of them, and then setting the expected penalty structure so that industry output does not exceed its CT-optimal regulated equilibrium level. In the case of the finite number of competitive firms analysed in the text, the marginal expected penalty determined the regulated equilibrium of the industry and the extent of illegal behaviour by member firms. Here the marginal expected penalty must be set at such a level that entry by non-quota holders is impossible, in the sense that any firm which does not hold a quota and attempts to enter incurs an unsustainable loss. Depending on the form of individual cost functions in the industry, this will require that the expected penalty function incorporate a flat-rate component. This type of penalty function will also preclude any constraint violation by quota holders, if set at such a level to prohibit entry, and thus allows the quota to be set at the regulated equilibrium output level which prevents any loss of industry profit through payment of fines.

This last point has implications for the analysis in the text. There too a CT regulator would, if able, choose to enforce an output quota with a flat unit rate expected penalty. A penalty of this type allows members of the industry, as quota holders, to capture, in the form of imputed quota rentals, all expected penalty payments which they would otherwise incur.

Finally, Section 3-8 briefly considered, within a similar analytical structure to that used in the preceding sections, the regulation of a monopolistic industry. A NPIT regulator was shown to devote fewer resources to restricting output in this instance than would be the case in regulating a competitive industry. The concept of strategic regulation, designed by private industry interests to preclude the possibility of more harmful NPIT measures, was considered. It was concluded that anything other than token control of the industry was, in these circumstances, an indication that the industry was not regulated by an unfettered CT regulator.

The following chapter presents an application of the analysis of Chapters Two and Three to the problem of regulating a common-property externality. This is done within the particular context of an open-access fishery.

NOTES.

1. Following (3-2-7) and (3-2-8) the parameter vector Z, which is used here consistent with the formulation of (3-2-5), comprises the state of enforcement technology $\mu$ and the level of consumer income Y.

2. See for example Harford, 1978; Lee, 1984; Papps, 1985; and Storey and McCabe, 1980.

3. Voluntarily is used here in the sense of a freely undertaken action. "Voluntary compliance" in the literature is often taken to mean compliance without direct enforcement. There are in those situations, however, other "costs" of noncompliance such as social stigma which act as effective deterrents. Here it is assumed that all such non-monetary influences are either irrelevant to behaviour or are quantified within the expected monetary penalty.

4. In the alternative case when $\partial Q/\partial \Omega = 0$, $T = 0$. From (3-3-5) then $dL_e^*/dY \gtrless 0$ if and only if $F \lessgtr 0$. The condition on the sign of F from (3-3-7) is therefore sufficient to produce the result in this case also.

5. The bounds $k_1$ and $k_2$ are not constants as is usually the case. Rather they are also variable containing a term in the variable that they bound. Substituting for S in (3-3-20) and (3-3-21) from (3-3-3) each of $k_1$, $k_2$ contain the term $[h(.) - SC_1(.)]\frac{\partial^2 Q}{\partial \Omega \partial \mu}$ and, following Lemma 3-2-1, the coefficient of this term is negative. Given that the Proposition considers regulation by sales tax only, the discussion in Appendix 3-2 shows that B and M are positive while

the concavity of (3-2-9) ensures that A and N are negative. Using this information in (3-3-20) and (3-3-21) shows that the coefficient on the term in $\partial^2 Q/\partial L_e \partial \mu$ is positive for both $k_1$ and $k_2$. The bounds $k_1$ and $k_2$ then vary directly with $\partial^2 Q/\partial L_e \partial \mu$ and provide a "tunnel" band within which the variable lies.

6. This is an extreme version of the "captured regulator" hypothesis. The constraints of the political process may limit the ability of the CT regulator to achieve its objective. See for instance Becker, 1983 ; and Peltzman 1976.

7. It has been suggested [ Waud, 1986; 213] that the existence of tax avoidance and tax evasion is a more convincing argument in support of the Laffer curve than the original which seeks to explain an inverse relationship between tax revenue and tax rates in terms of supply-side responses. (For an exposition of the original hypothesis see Fullerton, 1982).

8. This assumes that a change in the value of these parameters does not violate any overall profit constraint.

9. The text deals with zero declaration and partial declaration. Clearly a strictly positive sales tax with full declaration is not optimal for the CT regulator as this maximizes the loss in industry profits from tax payments at any given tax rate.

10. Taking any a, b, $0 \leq a, b \leq 1$ such that, following Lemma 3-7-2 part (i), $b \leq a$, and rewriting (3-7-1) in these general terms gives

$$b\omega'(L_e) + (1-a)\frac{\partial EP_{\Omega}(.)}{\partial L_e} > \omega'(L_e) \qquad (I)$$

and rearranging

$$b\omega'(L_e) - \frac{a\partial EP_{\Omega}(.)}{\partial L_e} > \omega'(L_e) - \frac{\partial EP_{\Omega}(.)}{\partial L_e} \qquad (II)$$

Replacing 'a' by 'b' in (II) and simplifying gives

$$b[\omega'(L_e) - \frac{\partial EP_{\Omega}(.)}{\partial L_e}] > \omega'(L_e) - \frac{\partial EP_{\Omega}(.)}{\partial L_e} \qquad (III)$$

Given the above assumption the left-hand side of (III) is at least as large as the left-hand side of (II) yet the inequality in (III) does not hold because by assumption $b \leq 1$. Given this, the assumed inequality in (II) does not hold either. The simplification employed in (3-7-1) therefore does not affect the generality of the result.

11. Here a finite number of identical forms are acting together to maximize joint profit. The profit-maximizing aggregate output level is then the same as that of a single owner of all the firms acting as a pure monopolist.

12. Alternatively it is possible that monopoly control of the industry will result in an output level which is less than is socially optimal. It may be that in these circumstances a NPIT regulator would seek to expand industry output by setting a quota below which output must not fall or by offering a subsidy. There is no reason to suspect that the theory is symmetric however. This offers an interesting extension to the analysis that is not attempted here. Recently, some work has been done on the problem of

regulating monopoly particularly from the perspective of designing incentive-compatible regulatory mechanisms which induce the monopolist to reveal information and also economise on the necessity for enforcement. See for example, Baron and Myerson, 1982; and Loeb and Magat, 1979.

13. The possibility of strategic regulation has been raised before [Salop and Schefman, 1983]. There regulation was used as a weapon by a dominant firm to inflate the costs of rivals and thus reduce competition.

14. This point again relates to the process within the political arena by which control of the regulatory agency is determined and which was alluded to in note 5.

15. This point again relates to the use of regulation as a strategic weapon which was mentioned in note 12.

APPENDIX 3-1

From (3-2-7) the equation for the inverse demand function, which shows the demand price at any regulated equilibrium quantity given fixed values of certain parameters is

$$(A3\text{-}1\text{-}1) \qquad P = h(Q(L_e,\Omega,\mu,Y),Y) \quad ; \quad h_1' < 0,\ h_2 > 0$$

The justifications for the signs of the partial derivatives are established in the text. It is evident that the demand price at any quantity will increase with a rise in income given the assumptions that the commodity is a normal good and that the price-elasticity of demand is non-zero. As a result of this shift in demand, however, the regulated equilibrium quantity will change and thus aggregate income (Y) is one of the component parameters of the parameter vector Z in (3-2-5).

Following (2-5-7) and (2-6-15), using the regulatory instrument proxy ($\Omega$) and parameterizing, at the regulated equilibrium

$$(A3\text{-}1\text{-}2) \qquad D(P(L_e,\Omega,\mu,Y),Y) = S(P(L_e,\Omega,\mu,Y),\mu)$$

Totally differentiating (A3-1-2) with $dL_e = d\Omega = d\mu = 0$ gives

$$(A3\text{-}1\text{-}3) \qquad \frac{\partial D(.)}{\partial P}\frac{\partial P}{\partial Y} + \frac{\partial D(.)}{\partial Y} - \frac{\partial S(.)}{\partial P}\frac{\partial P}{\partial Y} = 0$$

and rearranging

$$(A3\text{-}1\text{-}4) \qquad \frac{\partial P^*}{\partial Y} = \frac{-\ \partial D(.)/\partial Y}{\dfrac{\partial D(.)}{\partial P} - \dfrac{\partial S(.)}{\partial P}} \gtreqless 0 \text{ if and only if } \frac{\partial D(.)}{\partial Y} \gtreqless 0$$

where, following the assumptions of Chapter Two, $\partial D(.)/\partial P < 0$ and $\partial S(.)/\partial P > 0$.

Given the assumption that the commodity is a normal good, $\partial D(.)/\partial Y > 0$, and hence an increase in aggregate income raises market equilibrium price. The effect on equilibrium quantity is then found by examining the response of supply to the increase in price. Therefore

$$\text{(A3-1-5)} \quad \frac{\partial Q}{\partial Y} = \frac{\partial S(.)}{\partial P} \frac{\partial P^*}{\partial Y} \gtrless 0 \text{ if and only if } \frac{\partial P^*}{\partial Y} \gtrless 0$$

and, using the result from (A3-1-4) that $\partial P^*/\partial Y > 0$, it follows that $\partial Q/\partial Y > 0$. An increase in aggregate income therefore increases the regulated equilibrium output of the industry.

Totally differentiating (A3-1-2) with $dL_e = d\Omega = dY = 0$ gives

$$\text{(A3-1-6)} \quad \frac{\partial D(.)}{\partial P} \frac{\partial P}{\partial \mu} - \frac{\partial S(.)}{\partial P} \frac{\partial P}{\partial \mu} - \frac{\partial S(.)}{\partial \mu} = 0$$

and rearranging

$$\text{(A3-1-7)} \quad \frac{\partial P^*}{\partial \mu} = \frac{\partial S(.)/\partial \mu}{\frac{\partial D(.)}{\partial P} - \frac{\partial S(.)}{\partial P}} \gtrless 0 \text{ if and only if } \frac{\partial S(.)}{\partial \mu} \lessgtr 0$$

using the assumptions of Chapter Two as for (A3-1-4).

An improvement in the state of enforcement technology may be expected to increase the probability of detection, and thus the expected penalty, at any given level of enforcement. This is qualitatively equivalent to an increase in enforcement. Following Propositions 2-5-1 and 2-6-2 therefore, assuming that the increase in $\mu$ sufficiently increases the marginal expected penalty with respect to output, $\partial S(.)/\partial \mu < 0$ and thus, from (A3-1-7), an improvement in the state of enforcement technology raises the market equilibrium price. The effect on regulated equilibrium quantity is found by examining the response of demand to this change in price. Thus

$$(A3\text{-}1\text{-}8) \quad \frac{\partial Q}{\partial \mu} = \frac{\partial D(.)}{\partial P} \frac{\partial P^*}{\partial \mu} \overset{>}{\underset{<}{=}} 0 \text{ if and only if } \frac{\partial P^*}{\partial \mu} \overset{<}{\underset{>}{=}} 0$$

and, using the result from (A3-1-7) that $\partial P^*/\partial \mu > 0$, it follows that $\partial Q/\partial \mu < 0$. An improvement in the state of enforcement technology therefore reduces the regulated equilibrium output of the industry.

From (3-2-8) the equation for the enforcement-exclusive social cost of production at any regulated equilibrium quantity given values of the relevant parameters is

$$(A3\text{-}1\text{-}9) \quad SC = SC(Q(L_e, \Omega, \mu, Y), \phi) \; ; \; SC_1 > 0, \; SC_{11} > 0, \; SC_2 < 0, \; SC_{12} < 0$$

Marginal social cost of output at any given state of technology is positive while an increase in $\phi$, which represents an improvement in technology in that it reduces the extent of externality produced, lowers the social cost at any level of output produced in the industry. It is also assumed that improvements in this technology reduce the marginal social cost of output at any given output level. An example of this would be the implementation of pollution abatement technology which, while it may increase costs in the industry concerned, reduces the degree of externality which the industry generates and thus lowers the social cost of the industry's operation.

The regulated equilibrium, as shown in Chapter Two, is determined by individual competitive profit-maximizing behaviour within the regulatory environment. The parameter $\phi$ is assumed not to affect either demand or supply decisions. As such, changes in $\phi$ do not affect the regulated equilibrium and hence $\phi$ does not appear in the parameter vector

Z in (3-2-5).

Finally, as shown in (A3-1-8), the state of enforcement technology affects the regulated equilibrium output level of the industry. This parameter then also appears in the parameter vector Z in (3-2-5). Thus

$$\text{(A3-1-10)} \quad Q = Q(L_e,\Omega,\mu,Y) \; ; \; Q_3 < 0,\; Q_{13} < 0,\; Q_{22} \lesseqgtr 0$$

The result that $Q_3 < 0$ was established in (A3-1-8). The signs of this and of the other partial derivatives reflect the assumption that an improvement in the technology of enforcement increases deterrence at any combination of the policy instruments and also increases the marginal effectiveness of enforcement activity and the level of the regulatory instrument. In the case of a sales tax $Q_2 < 0$ and hence $Q_{23} < 0$ whereas in the case of an output quota $Q_2 > 0$ and therefore $Q_{23} > 0$. In both cases this reflects the assumption that the marginal deterrent capabilities of the regulatory instrument are enhanced by the introduction of improved enforcement technology.

Appendix 3-2

The assumption of concavity of the NPIT regulator's objective function given in (3-2-9) ensures that the determinant of the coefficient matrix $\begin{bmatrix} A & B \\ M & N \end{bmatrix}$ is positive and that individual components A and N are negative.

From (3-3-2)

$$\text{(A3-2-1)} \quad A = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e^2} + [h_1(.)\frac{\partial Q}{\partial L_e} - SC_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial L_e} - \omega''(L_e)$$

Following Lemma 3-2-1 $[h(.)-SC_1(.)]$ is negative and, from (3-2-2), it is assumed that the resource cost of enforcement is non-decreasing in $L_e$. The assumptions in (3-2-7) and (3-2-8) are that $h_1(.) < 0$ and that marginal social cost is strictly increasing in output while, from (3-2-1), it follows that $\partial Q/\partial L_e < 0$. Given these assumptions and results, a sufficient condition for $A < 0$ is that $\partial^2 Q/\partial L_e^2 > 0$. That is $A < 0$ if the marginal productivity of resources devoted to enforcement in reducing the regulated equilibrium output level of the industry declines as the level of enforcement rises.

From (3-3-3)

$$\text{(A3-3-2)} \quad N = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial \Omega^2} + [h_1(.)\frac{\partial Q}{\partial \Omega} - SC_{11}(.)\frac{\partial Q}{\partial \Omega}]\frac{\partial Q}{\partial \Omega}$$

Following the discussion of the component terms of (A3-2-1), and that, from (3-2-11), $\partial Q/\partial \Omega \geq 0$ at optimal quota levels while it is assumed that $\partial Q/\partial \Omega < 0$ at non-optimal tax rates, a sufficient condition for $N < 0$ is that $\partial^2 Q/\partial \Omega^2 > 0$. That is, $N < 0$ if the impact on the regulated industry equilibrium output level of successive increases in the tax rate, or successive reductions in available quota, at any given level of enforcement, is diminishing.

Given that A and N are negative then, in order for (3-3-2) and (3-3-3) to be satisfied, it is necessary in the case of regulation by a sales tax that terms B and M be positive and in the case of regulation by an output quota that they be negative. This follows from the observation that the regulator seeks to restrict the output of the industry and hence with a sales tax $dL_e$ and $d\Omega$ have the same sign while with an output quota they have opposite signs.

As B and M are the cross derivative terms, $\partial^2\pi/\partial L_e\partial\Omega$ and $\partial^2\pi/\partial\Omega\partial L_e$ respectively, and are thus identical, a condition on the sign of one suffices for both. From (3-3-3)

(A3-2-3) $$M = [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial\Omega\partial L_e} + [h_1(.)\frac{\partial Q}{\partial L_e} - SC_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial\Omega}$$

Following the discussion of the component terms of (A3-2-1) and that in the case of a sales tax it is assumed that $\partial Q/\partial L_e \leq 0$, a necessary condition for $M > 0$ is that $\partial^2 Q/\partial\Omega\partial L_e < 0$. That is B and M are positive only if the tax rate and the level of enforcement are mutually reinforcing deterrents.

In the case of an output quota assuming that $\partial Q/\partial\Omega > 0$, a necessary condition for $M < 0$ is that $\partial^2 Q/\partial\Omega\partial L_e > 0$. That is, recalling that the regulator controls output by reducing the amount of quota available to the industry, B and M are negative only if increases in the level of enforcement and reductions in the size of quota are mutually reinforcing deterrents.

Using (3-3-2) and (3-3-3) the size of the coefficients on the change in the enforcement technology parameter are now examined. From (3-3-2) the coefficient is

$$\text{(A3-2-4)} \qquad [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \mu} + \left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial \mu}$$

and is denoted by E in (3-3-14). Following the assumptions in (A3-1-10) and the results previously derived on other component terms of (A3-2-4), the first composite term of (A3-2-4) is positive while the second is negative. The overall sign of (A3-2-4) is therefore ambiguous.

Rearranging (A3-2-4), $E \gtrless 0$ if and only if

$$\text{(A3-2-5)} \qquad \frac{\partial^2 Q/\partial L_e \partial \mu}{\partial Q/\partial \mu} \gtrless - \frac{[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial L_e}}{[h(.)-SC_1(.)]}$$

given that $[h(.)-SC_1(.)]$, $\frac{\partial Q}{\partial \mu} < 0$.

From (3-3-3), the second coefficient is

$$\text{(A3-2-6)} \qquad [h(.)-SC_1(.)]\frac{\partial^2 Q}{\partial \Omega \partial \mu} + \left[[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial \Omega}\right]\frac{\partial Q}{\partial \mu}$$

and is denoted by S in (3-3-14). Following the assumptions in (A3-1-10) and the results previously derived on the other component terms of (A3-2-6), in the case of a sales tax, the first composite term of (A3-2-6) is positive and the second negative. The overall sign of S is therefore ambiguous also.

Rearranging (A3-2-6) for the case of a sales tax, $S \gtrless 0$ if and only if

$$\text{(A3-2-7)} \qquad \frac{\partial^2 Q/\partial \Omega \partial \mu}{\partial Q/\partial \mu} \gtrless - \frac{[h_1(.)-SC_{11}(.)]\frac{\partial Q}{\partial \Omega}}{[h(.)-SC_1(.)]}$$

given that $[h(.)-SC_1(.)]$, $\frac{\partial Q}{\partial \mu} < 0$.

Using the assumptions from (A3-1-10) in the case of an output quota, and that $\partial Q/\partial \Omega > 0$ following Proposition 2-6-2, the first term of (A3-2-6) is negative while the second term is positive. Rearranging (A3-2-6) then, $S \gtrless 0$ if and only if (A3-2-7) holds. Here, with an output quota, however, both sides of (A3-2-7) are negative whereas with a sales tax both are positive.

Appendix 3-3

The assumption of concavity of (3-4-16) ensures that the determinant of the coefficient matrix in (3-5-4) is positive and that individual components A and D are negative. From (3-5-1)

$$\text{(A3-3-1)} \quad A = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e{}^2} - \frac{\partial^2 w_R(.)}{\partial L_e{}^2} + \left[[h_{11}(Q(.),Y)Q(.) + h_1(Q(.),Y) + h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

Following Lemma 3-4-1 $[h_1(Q(.),Y) + h(Q(.),Y) - C'(Q(.))]$ is negative and, from (3-4-7), the marginal cost of enforcement to the industry is assumed to be non-decreasing. From (3-2-7) $h_1(.) < 0$ and, from (3-4-17), it follows that $\partial Q/\partial L_e < 0$ while from the concavity of the profit function $C''(Q(.)) > 0$. Given these assumptions and results, a sufficient condition for $A < 0$ is that $\partial^2 Q/\partial L_e{}^2 > 0$ and that

$$\text{(A3-3-2)} \qquad h_{11}(Q(.),Y) < C''(Q(.)) - 2h_1(Q(.),Y)$$

That is, $A < 0$ if the marginal productivity of resources devoted to enforcement in reducing the regulated equilibrium output level of the industry declines as the level of enforcement rises and the price responsiveness of the demand curve does not greatly increase as quantity rises.

From (3-5-2)

$$\text{(A3-3-3)} \quad N = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial R^2} - \frac{\partial^2 w_R(.)}{\partial R^2} + \left[[h_{11}(Q(.),Y)Q(.) + h_1(Q(.),Y) + h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial R}\right]\frac{\partial Q}{\partial R}$$

Following the discussion of the component terms of (A3-3-1) and that, from (3-4-18), $\partial Q/\partial R \geq 0$ at optimal quota levels, a sufficient condition for $N < 0$ is that $\partial^2 Q/\partial R^2 > 0$ and that $\partial^2 w_R(.)/\partial R^2 \geq 0$. That is, $N < 0$ if the impact on the regulated industry equilibrium output level of successive reductions in available quota, at any given level of enforcement, is decreasing, and the marginal cost to the industry of increasing the quota level is non-decreasing.

Given that A and N are negative then, in order for (3-5-1) and (3-5-2) to be satisfied, it is necessary that terms B and M be negative. This follows from the observation that with an output quota $dL_e$ and $dR$ have opposite signs. As B and M are the cross-derivative terms, a condition on the sign of one suffices for both. From (3-5-1)

$$\text{(A3-3-4)} \quad B = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial R} - \frac{\partial^2 w_R}{\partial L_e \partial R} + \Big| [h_{11}(Q(.),Y)Q(.) + h_1(Q(.),Y) + h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial R} \Big| \frac{\partial Q}{\partial L_e}$$

Following the discussion of the component terms of (A3-3-1) a necessary condition for $B < 0$ is that at least one of $\partial^2 Q/\partial L_e \partial R$, $\partial^2 w_R(.)/\partial L_e \partial R$ is positive. Using (3-4-5) and (3-4-7)

$$\text{(A3-3-5)} \quad \frac{\partial^2 w_R}{\partial L_e \partial R} = \frac{\partial^2 EP_R}{\partial L_e \partial R} = (1-a)[H_1 \frac{\partial^2 Q}{\partial L_e \partial R} + H_{12}\frac{\partial Q}{\partial L_e} + H_{32}] \ ;$$

$$H_{12}, H_{32} < 0$$

From the assumptions on component terms in (3-4-5) and assuming that an increase in quota reduces the marginal expected penalty with respect to enforcement, then, if

$\partial^2Q/\partial L_e\partial R > 0$, the sign of $\partial^2 w_R(.)/\partial L_e\partial R$ is ambiguous. This suggests that $B < 0$ only if $\partial^2Q/\partial L_e\partial R > 0$ which implies that increases in enforcement and reductions in quota are mutually reinforcing deterrents against illegal behaviour.

Assumptions must also be made about the components of the parameter coefficients in (3-5-1) and (3-5-2). From (3-4-7)

$$\text{(A3-3-6)} \qquad \frac{\partial^2 w_R}{\partial L_e \partial b} = \omega'(L_e) > 0$$

and

$$\text{(A3-3-7)} \qquad \frac{\partial^2 w_R}{\partial L_e \partial a} = \frac{\partial^2 EP_R}{\partial L_e \partial_a} = -[H_1 \frac{\partial Q}{\partial L_e} + H_3]$$

which is ambiguous in sign. Therefore

$$\text{(A3-3-8)} \qquad \frac{\partial^2 w_R}{\partial L_e \partial a} \lesseqgtr 0 \text{ if and only if } H_3 \gtreqless - H_1 \frac{\partial Q}{\partial L_e}$$

From (3-5-1) the coefficient on the change in the income parameter is

$$\text{(A3-3-9)} \quad F = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial Y}$$
$$- \frac{\partial^2 w_R}{\partial L_e \partial Y} + \Big[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial Y}$$
$$+ h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\Big]\frac{\partial Q}{\partial L_e}$$

Using (3-4-5) and (3-4-7)

$$\text{(A3-3-10)} \quad \frac{\partial^2 w_R}{\partial L_e \partial Y} = \frac{\partial^2 EP_R}{\partial L_e \partial Y} = (1-a)[H_{11}\frac{\partial Q}{\partial Y}\frac{\partial Q}{\partial L_e} + H_1 \frac{\partial^2 Q}{\partial L_e \partial Y}]$$

Following (A3-1-5), $\partial Q/\partial Y > 0$. Using this, the assumptions of (3-4-5), and that, from Section 2-6 $H_{11} > 0$, if $\partial^2Q/\partial L_e\partial Y = 0$, which is the case if the change in income does not affect the marginal productivity of enforcement, $\partial^2 w_R/\partial L_e\partial Y < 0$. If the change in income increases the marginal productivity of enforcement then this result is reinforced. If the opposite occurs the result may be reversed.

Following Lemma 3-4-1, the initial bracketed term is negative. If the change in income leaves the marginal productivity of enforcement unaltered, this term, when multiplied by $\partial^2 Q/\partial L_e \partial Y$, disappears. In this case, given that $\partial Q/\partial L_e < 0$, $F > 0$ if

$$\text{(A3-3-11)} \quad h_2(Q(.),Y) + h_{12}(Q(.),Y)Q(.) < \frac{\partial^2 w_R}{\partial L_e \partial Y} / \frac{\partial Q}{\partial L_e}$$

$$- [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial Y}$$

where, following (A3-3-2), the right-hand side of (A3-3-11) is positive. From (3-2-7), $h_2(Q(.),Y) > 0$. The inequality in (A3-3-11) therefore is more likely to hold the more negative is $h_{12}(Q(.),Y)$.

From (3-5-2) the coefficient on the change in the income parameter is

$$\text{(A3-3-12)} \quad T = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial R \partial Y}$$

$$- \frac{\partial^2 w_R}{\partial R \partial Y} + \Bigg[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial Y}$$

$$+ h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\Bigg]\frac{\partial Q}{\partial R}$$

From (3-4-20)

$$\text{(A3-3-13)} \quad \frac{\partial^2 w_R}{\partial R \partial Y} = \frac{\partial^2 EP_R}{\partial R \partial Y} = (1-a)[H_{11}\frac{\partial Q}{\partial Y}\frac{\partial Q}{\partial R} + H_1\frac{\partial^2 Q}{\partial R \partial Y}]$$

Given the assumptions of (3-4-20) and the discussion of (A3-3-10), if the change in income does not affect the marginal productivity of the quota, $\partial^2 w_R(.)/\partial R \partial Y > 0$. If the marginal productivity of the quota is enhanced by the increase in income, then this result is reinforced but, if the opposite occurs, the result may be reversed.

Assuming that the increase in income does not affect the marginal productivity of the quota, and given that $\partial Q/\partial R > 0$ then $T < 0$ if the inequality in condition (A3-3-11) holds.

From (3-5-1) the coefficient of the parameter representing the state of enforcement technology is

$$\text{(A3-3-14)} \quad E = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu}$$

$$- \frac{\partial^2 w_R}{\partial L_e \partial \mu} + \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y)\right.$$

$$\left. - C''(Q(.),Y)\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial L_e}$$

Rearranging (A3-3-14) and using Lemma 3-4-1, $E \gtreqless 0$ if and only if

$$\text{(A3-3-15)} \quad \frac{\partial^2 Q}{\partial L_e \partial \mu}$$

$$\lesseqgtr \frac{\dfrac{\partial w_R^2}{\partial L_e \partial \mu} - \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\dfrac{\partial Q}{\partial L_e}\right]\dfrac{\partial Q}{\partial \mu}}{[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]}$$

Following Lemma 3-4-1 the denominator of the right-hand side of (A3-3-15) is negative. Using (A3-3-2) and that, from (3-4-17), $\partial Q/\partial L_e < 0$, together with the assumption in (A3-1-10) that an improvement in the technology of enforcement increases deterrence at any combination of the policy instruments, it follows that the second term in the numerator of the right-hand side of (A3-3-15) is also negative. From (3-4-7)

$$\text{(A3-3-16)} \quad \frac{\partial^2 w_R}{\partial L_e \partial \mu} = \frac{\partial^2 EP_R}{\partial L_e \partial \mu}$$

Modifying (3-4-4) to include the parameter for the state of enforcement technology

$$\text{(A3-3-17)} \quad EP_R = (1-a)H(Q(L_e,R,\mu),R,L_e,\mu) \;;\; H_4 > 0,\; H_{24} < 0,\; H_{34} > 0$$

where it is assumed that an improvement in the state of enforcement technology increases the expected penalty faced at any level of illegal behaviour with given values of the policy instruments and also increases the absolute values of the marginal expected penalties with respect to enforcement and quota size at any given illegal output level. Using (3-4-5) in (A3-3-16)

$$\text{(A3-3-18)} \quad \frac{\partial^2 w_R}{\partial L_e \partial \mu} = (1-a)\left[H_1 \frac{\partial^2 Q}{\partial L_e \partial \mu} + H_{11} \frac{\partial Q}{\partial L_e} \frac{\partial Q}{\partial \mu} + H_{34}\right]$$

Given the assumption of convexity of the expected penalty function, $H_{11} > 0$, while, as in (A3-1-10), it is assumed that an improvement in the state of enforcement technology increases the marginal effectiveness of enforcement in reducing the extent of illegal activity so that $\partial^2 Q/\partial L_e \partial \mu < 0$. The sign of (A3-3-18) is therefore ambiguous.

If $\partial^2 w_R/\partial L_e \partial \mu > 0$ then the right-hand side of (A3-3-15) is negative. The more negative the term $\partial^2 Q/\partial L_e \partial \mu$ becomes, the less positive is $\partial w_R/\partial L_e \partial \mu$ using (A3-3-18) and hence the less negative is the right-hand side of (A3-3-15). Therefore, using (A3-3-15), $E > 0$ if $\partial^2 Q/\partial L_e \partial \mu$ is sufficiently negative. Here as in the NPIT case the relative sizes of the terms $\partial^2 Q/\partial L_e \partial \mu$ and $\partial Q/\partial \mu$ is also important. The term $\partial Q/\partial \mu$ reflects the magnitude of the improvement in deterrence at any given policy combination of enforcement and quota. An increase in the absolute magnitude of this term increases the positiveness of

$\partial^2 w_R / \partial L_e \partial \mu$ and hence increases the negativeness of the right-hand side of (A3-3-15).

From (3-5-2) the coefficient of the parameter reflecting the state of enforcement technology is

$$\text{(A3-3-19)} \quad [h_1(Q(.),Y) + h(Q(.),Y) - C'(Q(.))] \frac{\partial^2 Q}{\partial R \partial \mu} - \frac{\partial^2 w_R}{\partial R \partial \mu}$$

$$+ \left[ [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))] \frac{\partial Q}{\partial R} \right] \frac{\partial Q}{\partial \mu}$$

which is denoted by S. Rearranging (A3-3-19) using Lemma 3-4-1, $S \lesseqgtr 0$ if and only if

$$\text{(A3-3-20)} \quad \frac{\partial^2 Q}{\partial R \partial \mu}$$

$$\gtrless \frac{\frac{\partial^2 w_R}{\partial R \partial \mu} - \left[ [h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))] \frac{\partial Q}{\partial R} \right] \frac{\partial Q}{\partial \mu}}{[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]}$$

Following Lemma 3-4-1 the denominator of the right-hand side term of (A2-2-17) is negative. Using (A3-3-2) and assuming as above that $\partial Q / \partial \mu < 0$ while, following Proposition 2-6-2, $\partial Q / \partial R > 0$, the second term in the numerator of the right-hand side of (A3-3-20) is positive.

From (3-4-20) using (A3-3-17)

$$\text{(A3-3-21)} \quad \frac{\partial^2 w_R}{\partial R \partial \mu} = \frac{\partial^2 EP_R}{\partial R \partial \mu} = (1-a)[H_1 \frac{\partial^2 Q}{\partial R \partial \mu} + H_{11} \frac{\partial Q}{\partial R} \frac{\partial Q}{\partial \mu} + H_{24}]$$

Following (A3-1-10) it is assumed that improvement in the state of enforcement technology increases the marginal effectiveness of the quota in restricting output so that $\partial^2 Q / \partial R \partial \mu > 0$ and similarly, from (A3-3-17), the improved enforcement technology increases the absolute size of the marginal expected penalty with respect to quota size and hence $H_{24} < 0$. Given these assumptions, and the assumptions and conditions on the signs of other component terms in (A3-3-21), the sign of $\partial^2 w_R/\partial R\partial\mu$ is ambiguous.

If $\partial^2 w_R/\partial R\partial\mu < 0$ then the right-hand side of (A3-3-20) is positive. The more positive the term $\partial^2 Q/\partial L_e\partial\mu$ becomes, however, the less negative is $\partial^2 w_R/\partial R\partial\mu$, using (A3-3-21), and hence the less positive is the right-hand side of (A3-3-20). Therefore, using (A3-3-20), $S < 0$ if $\partial^2 Q/\partial L_e\partial\mu$ is sufficiently positive. As in (A3-3-15) the relationship between $\partial^2 Q/\partial L_e\partial\mu$ and $\partial Q/\partial\mu$ is important. An increase in the absolute magnitude of $\partial Q/\partial\mu$ increases the negativeness of $\partial^2 w_R/\partial R\partial\mu$ and hence increases the positiveness of the right-hand side of (A3-3-20).

Appendix 3-4

The assumption of concavity of (3-4-22) ensures that $\Delta > 0$ and that the individual components A and N are negative. From (3-5-24)

$$\text{(A3-4-1)}\quad A = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e{}^2} - \frac{\partial^2 w_t}{\partial L_e{}^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

Following Lemma 3-4-1 $[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]$ is negative and, from (3-4-12), the marginal cost of enforcement to the industry is assumed to be non-decreasing while from (3-4-23) it follows that $\partial Q/\partial L_e < 0$. Given these assumptions and results, a sufficient condition for $A < 0$ is that $\partial^2 Q/\partial L_e{}^2 > 0$ and that (A3-3-2) holds in the case of regulation by sales tax. That is $A < 0$ if the marginal effectiveness of resources devoted to enforcement in reducing the regulated equilibrium output of the industry declines as the level of enforcement rises at any given tax rate and the price responsiveness of the demand curve does not greatly increase as quantity rises.

From (3-5-25)

$$\text{(A3-4-2)}\quad N = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial t^2} - \frac{\partial^2 w_t}{\partial t^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial t}\right]\frac{\partial Q}{\partial t}$$

Following the discussion of the component terms of (A3-4-1) and that, from (3-4-24), $\partial Q/\partial t \leq 0$ at optimal tax rates, a sufficient condition for $N < 0$ is that $\partial^2 Q/\partial t^2 > 0$ and $\partial^2 w_t/\partial t^2 \geq 0$. That is $N < 0$ if the impact on the regulated equilibrium output level of successive increases in the tax rate at any given level of enforcement is decreasing, and the marginal cost to the industry of increasing the tax rate is non-decreasing as is assumed with reference to (3-4-25).

Given that A and N are negative, in order for (3-5-24) and (3-5-25) to be satisfied, it is necessary that B and M are positive. From (3-5-24)

$$\text{(A3-4-3)}\quad B = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial t}$$

$$- \frac{\partial^2 w_t}{\partial L_e \partial t} + \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial t}\right]\frac{\partial Q}{\partial L_e}$$

Following the discussion of the component terms of (A3-4-1) a necessary condition for $B > 0$ is that at least one of $\partial^2 Q/\partial L_e \partial t$, $\partial^2 w_t/\partial L_e \partial t$ is negative. Using (3-4-3), (3-4-9) and (3-4-12)

$$\text{(A3-4-4)}\quad \frac{\partial^2 w_t}{\partial L_e \partial t} = (1-a)[G_{11}\frac{\partial Q}{\partial L_e}\frac{\partial Q}{\partial t} + G_1\frac{\partial^2 Q}{\partial L_e \partial t} + G_{22}\frac{\partial X}{\partial t}\frac{\partial X}{\partial t} + G_2\frac{\partial^2 X}{\partial t^2}]$$

where $G_{11}, G_1, G_{22} > 0$, $G_2 < 0$.

The signs of the other component terms of (A3-4-4) are as previously established and in addition it is assumed that $\partial^2 X/\partial t^2 \leq 0$. If $\partial^2 Q/\partial L_e \partial t < 0$ the sign of $\partial^2 w_t/\partial L_e \partial t$ is ambiguous. This suggests that $B > 0$ only if $\partial^2 Q/\partial L_e \partial t < 0$ which implies that increases in enforcement and increases in the tax rate are mutually reinforcing deterrents against illegal behaviour.

Assumptions must also be made about the components of the parameter coefficients in (3-5-24) and (3-5-25). Using (3-4-12)

$$\text{(A3-4-5)}\qquad \frac{\partial^2 w_t}{\partial L_e \partial b} = \omega'(L_e) > 0$$

and using (3-4-12) and (3-4-3)

$$\text{(A3-4-6)} \quad \frac{\partial^2 w_t}{\partial L_e \partial a} = -[G_1 \frac{\partial Q}{\partial L_e} + G_2 \frac{\partial X}{\partial L_e} + G_3] - \frac{\partial \tau}{\partial L_e}$$

From (3-4-9) $\partial\tau/\partial L_e$ is non-negative so that,using (3-4-3), $\partial^2 w_t/\partial L_e \partial a < 0$ if $\partial EP_t/\partial L_e > 0$ and $\partial^2 w_t/\partial L_e \partial a > 0$ only if the marginal expected fine payments which result from an increase in enforcement are sufficiently negative.

From (3-4-25) and (3-4-26)

$$\text{(A3-4-7)} \quad \frac{\partial^2 w_t}{\partial t \partial a} = -[G_1 \frac{\partial Q}{\partial t} + G_2 \frac{\partial X}{\partial t}] - [\tau_1 \frac{\partial X}{\partial t} + \tau_2]$$

Using the assumptions in the text that $\partial EP_t/\partial t > 0$ and that this outweighs any negativity in $\partial\tau/\partial t$, $\partial^2 w_t/\partial t \partial a < 0$.

From (3-5-24) the coefficient on the change in income parameter is

$$\text{(A3-4-8)} \quad F = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial Y} - \frac{\partial^2 w_t}{\partial L_e \partial Y}$$
$$+ \Big[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial Y}$$
$$+ h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\Big]\frac{\partial Q}{\partial L_e}$$

From (3-4-12)

$$\text{(A3-4-9)} \quad \frac{\partial^2 w_t}{\partial L_e \partial Y} = \frac{\partial^2 EP_t}{\partial L_e \partial Y} + (1-a)\frac{\partial^2 \tau}{\partial L_e \partial Y}$$

Using (3-4-3) and (3-4-9)

$$\text{(A3-4-10)} \quad \frac{\partial^2 w_t}{\partial L_e \partial Y} = (1-a)[G_{11} \frac{\partial Q}{\partial L_e} \frac{\partial Q}{\partial Y} + G_1 \frac{\partial^2 Q}{\partial L_e \partial Y} + G_{22} \frac{\partial X}{\partial L_e} \frac{\partial X}{\partial Y}$$
$$+ G_2 \frac{\partial^2 X}{\partial L_e \partial Y}] + (1-a)[\tau_{11} \frac{\partial X}{\partial L_e} \frac{\partial X}{\partial Y} + \tau_1 \frac{\partial^2 X}{\partial L_e \partial Y}]$$

From previously derived results and assumptions, $G_1, G_{11}, G_{22}, \tau_1, \frac{\partial Q}{\partial Y}, \frac{\partial X}{\partial L_e} > 0$ while $\frac{\partial Q}{\partial L_e}, G_2 < 0$ and $\tau_{11} = 0$.

If the change in income does not affect the productivity of the regulatory instruments then $\partial^2 Q/\partial L_e \partial Y = 0$ and $\partial^2 X/\partial L_e \partial Y = 0$. In this case if $\partial X/\partial Y < 0$ then $\partial^2 w_t/\partial L_e \partial Y < 0$ but if $\partial X/\partial Y > 0$ the sign of $\partial^2 w_t/\partial L_e \partial Y$ is ambiguous. If, however, the change in income does affect the productivity of the regulatory instruments, the sign of $\partial^2 w_t/\partial L_e \partial Y$ is ambiguous irrespective of whether the effect is to improve instrument productivity or reduce it.

From (3-5-25) the coefficient of the change in income parameter is

$$\text{(A3-4-11)}\quad T = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial t \partial Y} - \frac{\partial^2 w_t}{\partial t \partial Y} + \Big[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.) + h_2(Q(.),Y)\Big]\frac{\partial Q}{\partial t}$$

Using (3-4-25) and (3-4-26)

$$\text{(A3-4-12)}\quad \frac{\partial^2 w_t}{\partial t \partial Y} = (1-a)[G_{11}\frac{\partial Q}{\partial t}\frac{\partial Q}{\partial Y} + G_1\frac{\partial^2 Q}{\partial t \partial Y} + G_{22}\frac{\partial X}{\partial t}\frac{\partial X}{\partial Y} + G_2\frac{\partial^2 X}{\partial t \partial Y}] + (1-a)[\tau_{11}\frac{\partial X}{\partial t}\frac{\partial X}{\partial Y} + \tau_2\frac{\partial^2 X}{\partial t \partial Y}]$$

The component terms of the right-hand side of (A3-4-12) are qualitatively identical to those of (A3-4-10) and hence the conclusions concerning the sign of $\partial^2 w_t/\partial t \partial Y$ are the same as for $\partial^2 w_t/\partial L_e \partial Y$.

From (3-5-24) the coefficient on the parameter representing the state of enforcement technology is

$$\text{(A3-4-13)}\quad E = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial L_e \partial \mu} - \frac{\partial^2 w_t}{\partial L_e \partial \mu} + \Big[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial \mu}\Big]\frac{\partial Q}{\partial L_e}$$

Rearranging (A3-4-13) and using Lemma 3-4-1, E $\gtrless$ 0 if and only if

$$(\text{A3-4-14}) \quad \frac{\partial^2 Q}{\partial L_e \partial \mu}$$

$$\lesseqgtr \frac{\dfrac{\partial^2 w_t}{\partial L_e \partial \mu} - \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\dfrac{\partial Q}{\partial L_e}\right]\dfrac{\partial Q}{\partial \mu}}{[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]}$$

Following Lemma 3-4-1, the denominator of the right-hand side of (A3-4-14) is negative. Using (A3-3-2) and that, from (3-4-23), $\partial Q/\partial L_e < 0$, together with the assumption in (A3-1-8) that $\partial Q/\partial \mu < 0$, it follows that the second term in the numerator of the right-hand side of (A3-4-14) is negative. From (3-4-12) using (3-4-3) and (3-4-9)

$$(\text{A3-4-15}) \quad \frac{\partial^2 w_t}{\partial L_e \partial \mu} = \frac{\partial^2 EP_t}{\partial L_e \partial \mu} + (1-a)\frac{\partial^2 \tau}{\partial L_e \partial \mu}$$

$$= (1-a)[G_{11}\frac{\partial Q}{\partial L_e}\frac{\partial Q}{\partial \mu} + G_1\frac{\partial^2 Q}{\partial L_e \partial \mu} + G_{22}\frac{\partial X}{\partial L_e}\frac{\partial X}{\partial \mu} + G_2\frac{\partial^2 X}{\partial L_e \partial \mu} + G_{34}]$$

$$+ (1-a)[\tau_{11}\frac{\partial X}{\partial L_e}\frac{\partial X}{\partial \mu} + \tau_1\frac{\partial^2 X}{\partial L_e \partial \mu}]$$

The assumptions on the component terms of the right-hand side of (A3-4-15) are that an improvement in the state of enforcement technology increases the expected penalty faced at any illegal behaviour level with given values of the policy instruments, increases the absolute values of the marginal expected penalties with respect to enforcement level, and tax rate at any given illegal output level and encourages more truthful tax declarations. Thus $\frac{\partial X}{\partial \mu}, \frac{\partial^2 X}{\partial L_e \partial \mu}, G_{34} > 0$. As before the assumption that $\tau_{11} = 0$ reflects the fact that tax revenue increases at a constant rate with declared output at any given tax rate. Given the assumption that an improvement in the state of enforcement technology improves the marginal effectiveness of enforcement in restricting industry output, so that $\partial^2 Q/\partial L_e \partial \mu < 0$, the sign of (A3-4-15) is ambiguous.

If $\partial^2 w_t/\partial L_e \partial \mu > 0$ then the right-hand side of (A3-4-14) is negative. The more negative the term $\partial^2 Q/\partial L_e \partial \mu$ becomes, the less positive is $\partial^2 w_t/\partial L_e \partial \mu$ and hence the less negative is the right-hand side of (A3-4-14). Therefore, from (A3-4-14), $E > 0$ if $\partial^2 Q/\partial L_e \partial \mu$ is sufficiently negative. Here as in the CT case of an output quota the relative sizes of the terms $\partial^2 Q/\partial L_e \partial \mu$ and $\partial Q/\partial \mu$ are also important. An increase in the absolute magnitude of $\partial Q/\partial \mu$, which reflects the improvement in deterrence at any given policy combination of enforcement and tax rate, increases the positiveness of $\partial^2 w_R/\partial L_e \partial \mu$ and the negativeness of the second term in the numerator of the right-hand side of (A3-4-14), and hence increases the overall negativeness of the right-hand side of (A3-4-14).

From (3-5-25) the coefficient of the parameter reflecting the state of enforcement technology is

$$\text{(A3-4-16)} \quad S = [h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]\frac{\partial^2 Q}{\partial t \partial \mu} - \frac{\partial^2 w_t}{\partial t \partial \mu} + \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\frac{\partial Q}{\partial \mu}\right]\frac{\partial Q}{\partial t}$$

Rearranging (A3-4-16) using Lemma 3-4-1, $S \gtrless 0$ if and only if

$$\text{(A3-4-17)} \quad \frac{\partial^2 Q}{\partial t \partial \mu} \lessgtr \frac{\dfrac{\partial^2 w_t}{\partial t \partial \mu} - \left[[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C''(Q(.))]\dfrac{\partial Q}{\partial t}\right]\dfrac{\partial Q}{\partial \mu}}{[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C'(Q(.))]}$$

Following Lemma 3-4-1 the denominator of the right-hand side of (A3-4-17) is negative. Using (A3-3-2) and assuming as above that $\partial Q/\partial\mu < 0$ while, following Proposition 2-5-1 and (3-4-24), $\partial Q/\partial t < 0$, the second term in the numerator of the right-hand side of (A2-3-17) is negative.

Using (3-4-25) and (3-4-26)

$$\text{(A3-4-18)} \qquad \frac{\partial^2 w_t}{\partial t \partial\mu} = \frac{\partial^2 EP_t}{\partial t \partial\mu} + (1-a)\frac{\partial^2 \tau}{\partial t \partial\mu}$$

$$= (1-a)\left[G_{11}\frac{\partial Q}{\partial t}\frac{\partial Q}{\partial\mu} + G_1\frac{\partial^2 Q}{\partial t\partial\mu} + G_{22}\frac{\partial X}{\partial t}\frac{\partial X}{\partial\mu} + G_2\frac{\partial^2 X}{\partial t\partial\mu}\right]$$

$$+ (1-a)\left[\tau_{11}\frac{\partial X}{\partial t}\frac{\partial X}{\partial\mu} + \tau_1\frac{\partial^2 X}{\partial t\partial\mu}\right]$$

Given the signs on the component terms of (A3-4-18) previously derived and the assumption that an improvement in the state of technology increases the marginal effectiveness of the tax rate in restricting output so that $\partial^2 Q/\partial t\partial\mu < 0$, the overall sign of (A3-4-18) is also ambiguous.

If $\partial^2 w_t/\partial t\partial\mu > 0$ then the right-hand side of (A3-4-17) is negative. The more negative the term $\partial^2 Q/\partial t\partial\mu$ becomes, the less positive is $\partial^2 w_t/\partial t\partial\mu$ and hence the less negative is the right-hand side of (A3-4-17). Therefore, using (A3-4-17), $S > 0$ if $\partial^2 Q/\partial t\partial\mu$ is sufficiently negative. As in previous cases the relative sizes of $\partial^2 Q/\partial t\partial\mu$ and $\partial Q/\partial\mu$ are also important. An increase in the absolute magnitude of $\partial Q/\partial\mu$ increases the negativeness of the right-hand side of (A3-4-17) and thus reduces the likelihood that $S > 0$ for any given value of $\partial^2 Q/\partial t\partial\mu$.

# CHAPTER FOUR

## REGULATION OF A COMMON PROPERTY RESOURCE: THE OPEN-ACCESS FISHERY*

### 4-1 INTRODUCTION

In Chapter Two, a partial equilibrium model of regulation for the case of a competitive negative-externality generating industry was derived. The concept of a regulated equilibrium was defined and illustrated using the examples of an output quota and a unit rate sales tax. In each case the regulatory instrument was enforced by means of an expected monetary penalty. The analysis also examined the effects on a regulated equilibrium of changes in the size of the regulatory instrument and in the level of enforcement. In Chapter Three the characteristics of the regulated equilibrium that emerges from all possible regulated equilibria, which comprise the form and level of the regulatory instrument used, the regulated equilibrium output level, and the amount of enforcement, were shown to depend on the objectives of the regulator. In principle then, it is possible to infer a regulator's objectives from the observable characteristics of a regulated equilibrium. It was shown that the most clearly differentiated behaviour emerges over the use of output quota as a regulatory instrument. Here the model is applied to the problem of the regulatory control of a common property externality.

The common property externality is a long recognised phenomenon in economics and has been frequently cited as one of the prime cases of apparent market failure. A widely used example of this phenomenon, particularly in the historical literature, is that of the open-field agricultural system found in pre-feudal Europe. This is the traditional "tragedy of the commons" problem in which, because private returns to herd expansion exceed private costs, but not social costs, overgrazing occurs [Hardin, 1968; 1244].

Several writers have questioned the appropriateness of the open-field or common as an example of market failure. Dahlman [1980] states that the commons were mainly restricted to non-arable land over which the community as a whole exercised usage rights and excluded entry by outsiders. Runge [1981] argues that this communal nature of resource ownership affects the inferences that can be drawn and contrasts it with the implications for resource use and allocation which result from open-access.

Under joint ownership and control, individual choices are influenced by expectations of the actions of other agents and thus the externality is nonseparable [Runge, 1981; 599]. With a non-separable externality, individual marginal cost functions are affected by the actions of all other agents so that the game theoretic Pareto inferior equilibrium, which characterises the prisoners' dilemma where externalities are separable, is possible but no longer the dominant strategy. He then suggests that common property externalities may be viewed as an assurance problem with interdependent choices forming a cooperative game where "it is in the interest of each (agent) to restrict output .... if that is the only way to get other agents to do likewise." [Runge, 1981; 600]. From this approach the operation of commercial agriculture is identical in structure to a cartel. The cooperative nature of the game in no way obviates the need for internal enforcement as each individual agent still has an incentive to overgraze unless the private cost to him of so doing is greater than the private benefit.

Dahlman [1980] in analysing the emergence of the open-field system within a property rights framework alludes to this point. A property right is a method of internalizing an externality and confers on its holder the ability to exercise certain rights over a resource, the most important of which being the right to exclude. As Coase [1960] originally argued, externalities can be accounted for by the market provided that the transactions costs of doing so are not prohibitive. Installing and maintaining a regime of property rights incurs these transactions costs and thus may not occur under market conditions. Dahlman suggests that transactions costs in pre-feudal non-arable agriculture were such as to facilitate the emergence and retention of communal property rights over open fields and prohibit the formation of private property rights.

If transactions costs are such as to prevent any form of property right over a resource,or the community of agents is arbitrarily large, then there is said to exist open-access to the resource in the context of unrestricted entry to resource use. In this case, the externality is separable, individual agent strategy is strictly dominant and,as Cheung [1970] derives, over-exploitation of the resource, at least in the economic sense, is inevitable. It is with this open-access situation, where resource-use and allocation inefficiencies exist, that the present chapter is concerned. The specific example used is that of an open-access fishery. Before the characteristics of open-access equilibrium are derived in Section 4-3, a brief review of the literature on the economics of fisheries is presented in the following section.

## 4-2 REVIEW OF THE LITERATURE

Much has been written in the area of fishery economics stemming mainly from the seminal paper by Gordon [1954] in which he illustrates the intramarginal dissipation of rent between fishing grounds and presents a model of the 'bionomics' of a fishery. The literature separates into two broad approaches. Gordon's work, and that following directly from it, employs a steady-state analysis in which the dynamics of the emergence of, and transition between, various equilibria are not explicitly considered. The papers which stem from Scott [1955], however, emphasize these aspects of the problem, examining the implications of time, and elaborate models of population dynamics, for the optimal management of the fishery.

Christy and Scott [1965] reformulate Gordon's analysis to show diagrammatically the economic interpretation of overfishing and give comparative static results of parameter changes. They also show that the biological concept of maximum sustainable yield (MSY) is not an economically efficient point for the fishery to operate at as the marginal return to input factors at this point is zero.

Smith [1966] however criticises this result for being dependent on the assumptions made about costs and argues that the form of the cost and revenue functions is not consistent with the optimizations used. In particular he shows that with a cost of extraction function independent of population size, which he asserts is the author's assumption, the optimum for a sole owner occurs at the MSY while with a fixed output price, free entry ensures that, if fishing is profitable at all, entry occurs until the fishery is exhausted. These results do not arise if the cost function is modified to allow for the likelihood that the costs of extraction increase as the population declines.

Turvey [1964] also uses steady-state analysis but relaxes the assumption of perfectly elastic demand employed in the other models and argues that the optimal harvest amount must account for consumer surplus in addition to rent yielded by the resource. Copes [1970] extends this analysis and derives a long-run supply curve for the fishery from factor cost and stock/yield relationships. This allows the problem to be formulated in familiar supply/demand terms and provides a useful framework in which to consider regulatory problems and objectives.

The majority of these analyses are conducted at the industry level but Anderson [1976] demonstrates the incentive linkages that exist for the individual firm and how their responses are incorporated into the industry results. Anderson [1977] also provides a broad overview of this branch of the literature producing further comparative static results within the Christy and Scott formulation and discussing the implications of time and discounting for the static optima.

Consideration of time allows for a more explicit examination of the dynamics of the system. Scott [1955] in an early revision of Gordon includes a discussion on the implications of short-run and long-run time horizons for decisions made about fishery management. In the long-run case, the optimal policy objective is shown to be that output level which maximizes the net present value of the stream of factor returns to the fishery given the appropriate discount rate.

Current extraction affects current stock levels, population growth, future stock levels and hence future extraction possibilities. Viewed in this manner, the fish population is analogous to a capital stock and thus capital theory may be applied to the analysis of optimal extraction. A reduction in current harvest which boosts population growth is then an investment activity.

These models [Clark and Munro, 1982; Quirk and Smith, 1970] employ sophisticated optimal control and dynamic programming techniques. In a simple linear control model, optimal extraction is characterized by the equivalent of the familiar "golden rule" where the own rate of return to the fish stock, consisting of its marginal physical product at the current stock level together with the marginal stock effect of harvest, equates with the social discount rate [Clark and Munro; 37-38]. They must also of necessity consider the complicated dynamics of the stock/yield relationships. Smith [1968, 1969] illustrates these properties giving conditions for the stability of the system and for the emergence of steady-state sustainable equilibria.

For the purposes of the present analysis, the fishery is assumed to exhibit stability and thus the steady-state approach is used. While dynamic properties are important in a model of renewable natural resources, the steady-state assumptions simplify the analysis and facilitate a clear comparison of differing regulatory objectives.

Leaving aside complications such as "mesh" externalities caused by different net mesh sizes affecting the age structure of the population and "crowding" externalities resulting from vessel congestion in a concentrated fishing ground [Smith, 1969; 181], the essential feature of the fishery is the stock externality which results from the stock/yield relationship. A fish taken at the margin alters stock size present and future, as demonstrated below, causing an external diseconomy to other input factors involved in the industry. This is the essence of the traditional tragedy of the commons argument and reflects a divergence between individual and collective marginal costs.

The following section presents the simple population dynamics of a fishery as contained in the literature cited above. These, when combined with demand conditions and production costs, generate the unregulated competitive equilibrium in the open-access fishery. This equilibrium is derived and its properties examined.

## 4-3 EQUILIBRIUM IN AN OPEN-ACCESS FISHERY

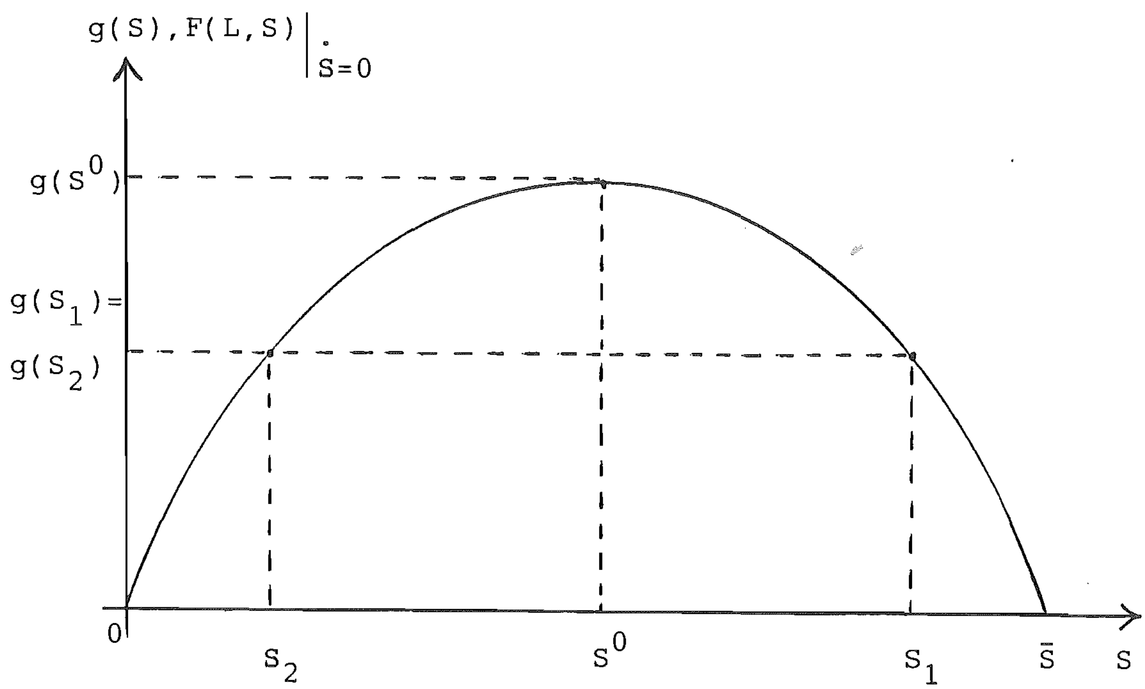
Within the fishery at any point in time there is a fish stock or population size denoted by S. $\dot{S} = dS/dt$ is the change in stock size over time. As with any natural population, additions to the stock occur through births while reductions follow from mortality. The difference between these two figures gives the net growth in population (G) which is assumed to be a function of stock size with the following properties:

$$G = g(S) \quad ; \quad g(0) = g(\bar{S}) = 0, \; \bar{S} > 0; \tag{4-3-1}$$
$$g'(0) > 0, \; g'(\bar{S}) < 0, \; g''(S) < 0$$

The possibility of a positive minimum population survivability threshold is precluded by the assumption $g(0) = 0$. There is however a maximum biologically sustainable population $\bar{S} > 0$ such that $g(\bar{S}) = 0$ also. For positive stock levels between these two bounds, growth is positive and variable. At small stock levels, net population growth increases with stock size while at larger levels, growth is positive but decreasing. This is because mortality rises more quickly than births [Smith, 1966; 1342] and therefore $g''(S) < 0$. Consequently, there is a certain stock level $S^0$ such that net population growth is maximized and $g'(S^0) = 0$. This is the population level that biologists refer to as that of maximum sustainable yield.

The growth function then has an inverted 'U' shape as shown in Figure 4-3-1 below.

Figure 4-3-1: Sustainable yield curve




This function g(S) gives the change in population under natural conditions. The introduction of fishing activity,however,also impacts on stock size. The economy is assumed to have labour as its single variable factor of production. The amount of fish caught (X) is a function of labour employed in the fishing industry and the size of the fish stock. Thus

(4-3-2) $X = F(L,S)$ ; $F_L > 0$, $F_{LL} < 0$, $F_S > 0$ ; $F(L,0) = 0$

The change in the fish population in any period is given by the difference between the natural population growth and the fish harvest in that period.

(4-3-3) $\dot{S} = g(S) - F(L,S)$

As mentioned above, the analysis here is confined to the steady-state case where $\dot{S} = 0$ and the stock size is constant over time. In the steady-state therefore, using (4-3-3)

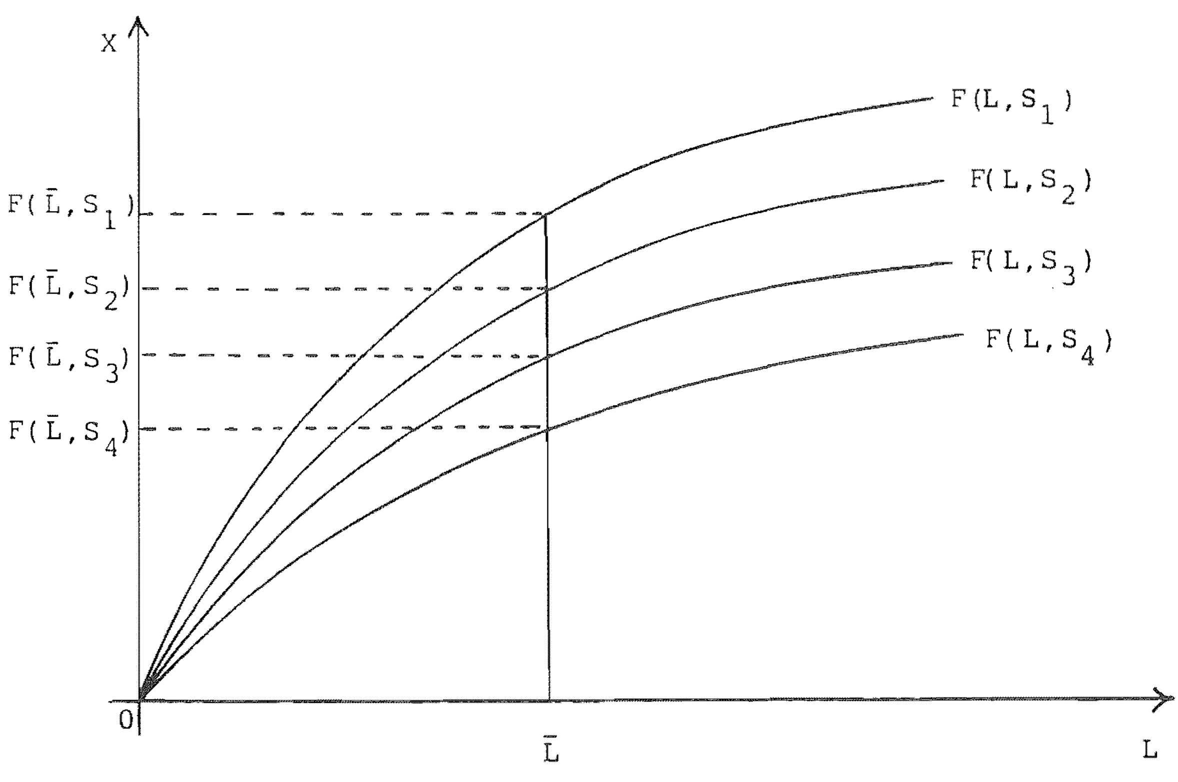
$$(4\text{-}3\text{-}4) \qquad F(L,S)\Big|_{\dot{S} = 0} = g(S) = \bar{X}(S)$$

The steady-state catch as a function of stock size then is also shown by the inverted 'U' curve in Figure 4-3-1 above.

The catch function as defined in (4-3-2) exhibits diminishing returns to labour, given a fixed stock level, and is affected parametrically by changes in stock size. Hence for a given amount of labour, $\bar{L}$, an increase in the fish stock increases the size of the technologically feasible catch. This is illustrated in Figure 4-3-2 where

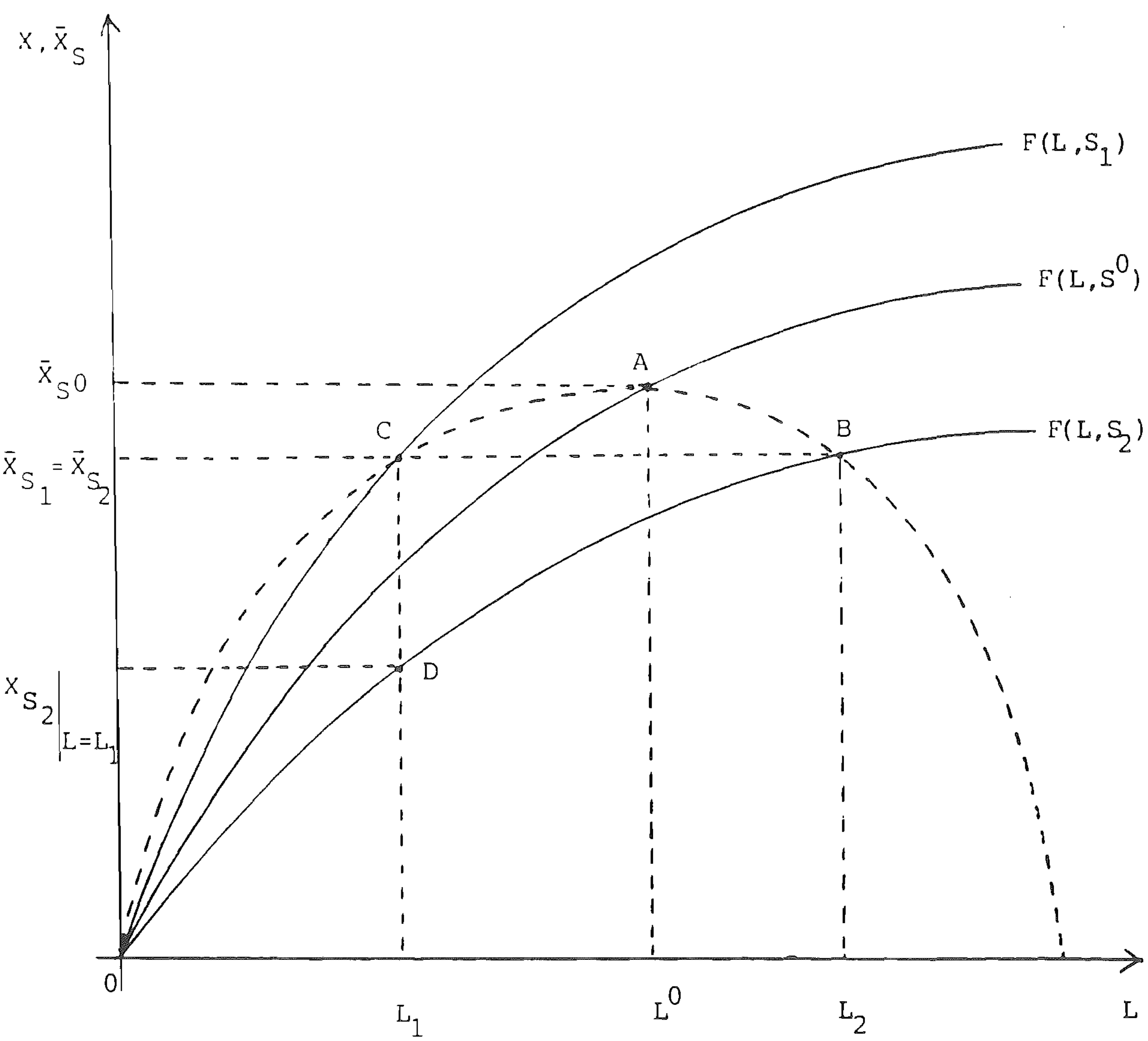
$$(4\text{-}3\text{-}4) \qquad F(\bar{L},S_i) > F(\bar{L}, S_{i+1}) \quad , \quad i = 1,2,3;\ S_i > S_{i+1}$$

Figure 4-3-2: Parametric changes in stock size



The catch functions in Figure 4-3-2 illustrate technical possibilities but do not account for biological feasibility. They are short-run production functions which can be related to sustainable yields through the information contained in Figure 4-3-1. Each of the stock sizes, $S_1,\ldots,S_4$, has a corresponding sustainable yield of amount $g(S_1),\ldots.g(S_4)$. Figures 4-3-1 and 4-3-2 can then be combined to give a relationship between sustainable yield and the labour employed in the fishery shown in Figure 4-3-3 below.

Figure 4-3-3: Sustainable Harvest Function



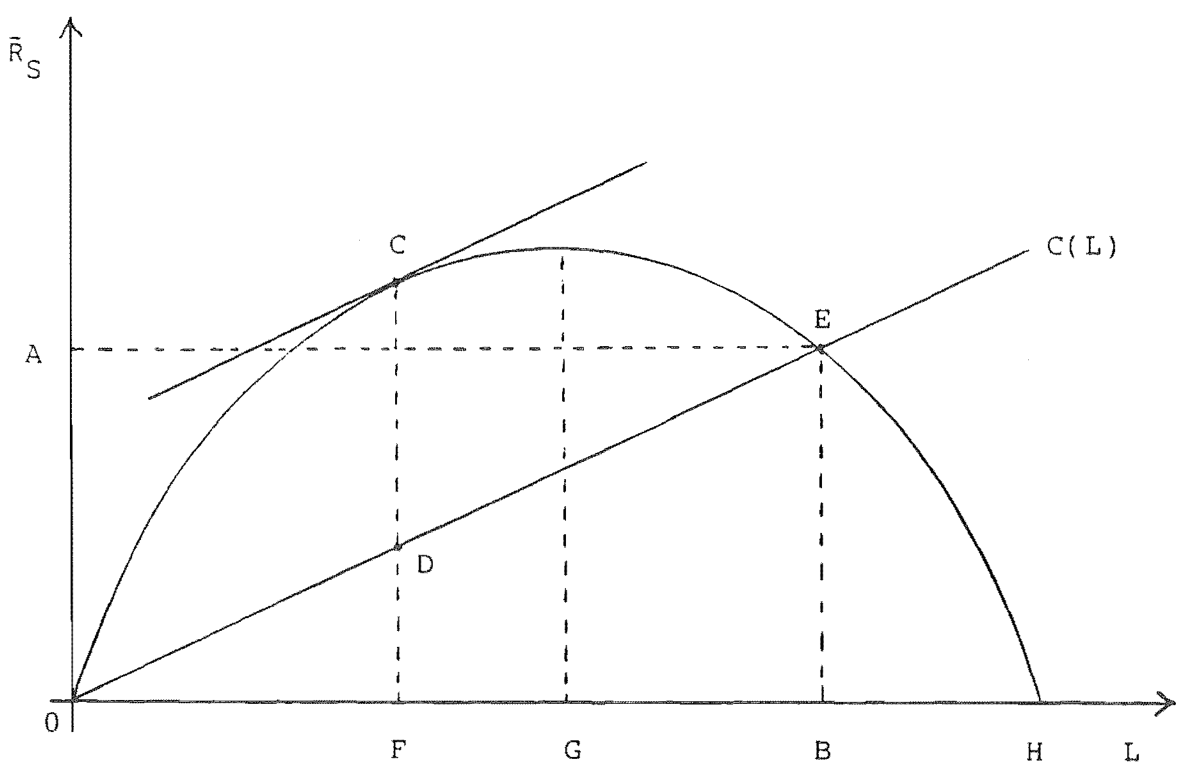
The production function $F(L,S^0)$ in Figure 4-3-3 corresponds to the MSY stock size. Following Figure 4-3-1 and using (4-3-4) there is a steady-state harvest size $\overline{X}_{S^0}$ associated with this stock level. Point 'A' then shows the technically efficient labour input, $L^0$, required for the steady-state catch level $\overline{X}_{S^0} = g(S^0)$. Figure 4-3-1 also demonstrates that there are two stock sizes which generate any smaller, strictly positive, sustainable yield. For example, stock sizes $S_1$ and $S_2$ both produce a steady-state catch of $\overline{X}_{S_1} = \overline{X}_{S_2}$ with $S_1$ and $S_2$. Given the respective short-run production functions, however, these catches are technically consistent with different factor input levels $L_1 < L_2$. This is shown by points B and C in Figure 4-3-3.

The dynamics of the relationship between short-run catches and steady-state yields can also be seen from this diagram. From an initial steady-state equilibrium at point B, with stock level $S_2$ and labour input $L_2$, assume that labour input to the fishery is reduced to level $L_1$. From the short-run production function, this implies a smaller catch size associated with point D. The reduced catch allows greater net population growth and increases stock levels until point C is attained where the stock size $S_1$ is such as to produce a sustained catch level technically consistent with the labour input $L_1$.[1]

Much analysis has been based on this sustainable yield curve 0CAB or its revenue equivalent. Under the fixed demand price assumption the sustainable revenue curve has the same shape as that in Figure 4-3-3 [Christy and Scott, 1965; 7]. With a negatively sloped linear demand curve, its shape depends on whether the maximum sustainable yield $\bar{X}_{S^0}$ is greater than or less than the quantity at which demand is unitary price elastic. If greater than, the possibility of multiple equilibria further complicates the results [Anderson, 1977; 82].

This sustainable revenue curve, when combined with the industry cost curve, which is usually taken to be linear by the assumption of constant long-run marginal cost, shows that open-access leads to over-exploitation of the resource. This overfishing may be in the economic sense, where productive factors are invested to the extent that zero resource rent is yielded, and/or in the biological sense with respect to the maximum sustainable yield [Christy and Scott, 1965; 8]. Figure 4-3-4 below illustrates the situation.

Figure 4-3-4: Christy and Scott open-access equilibrium



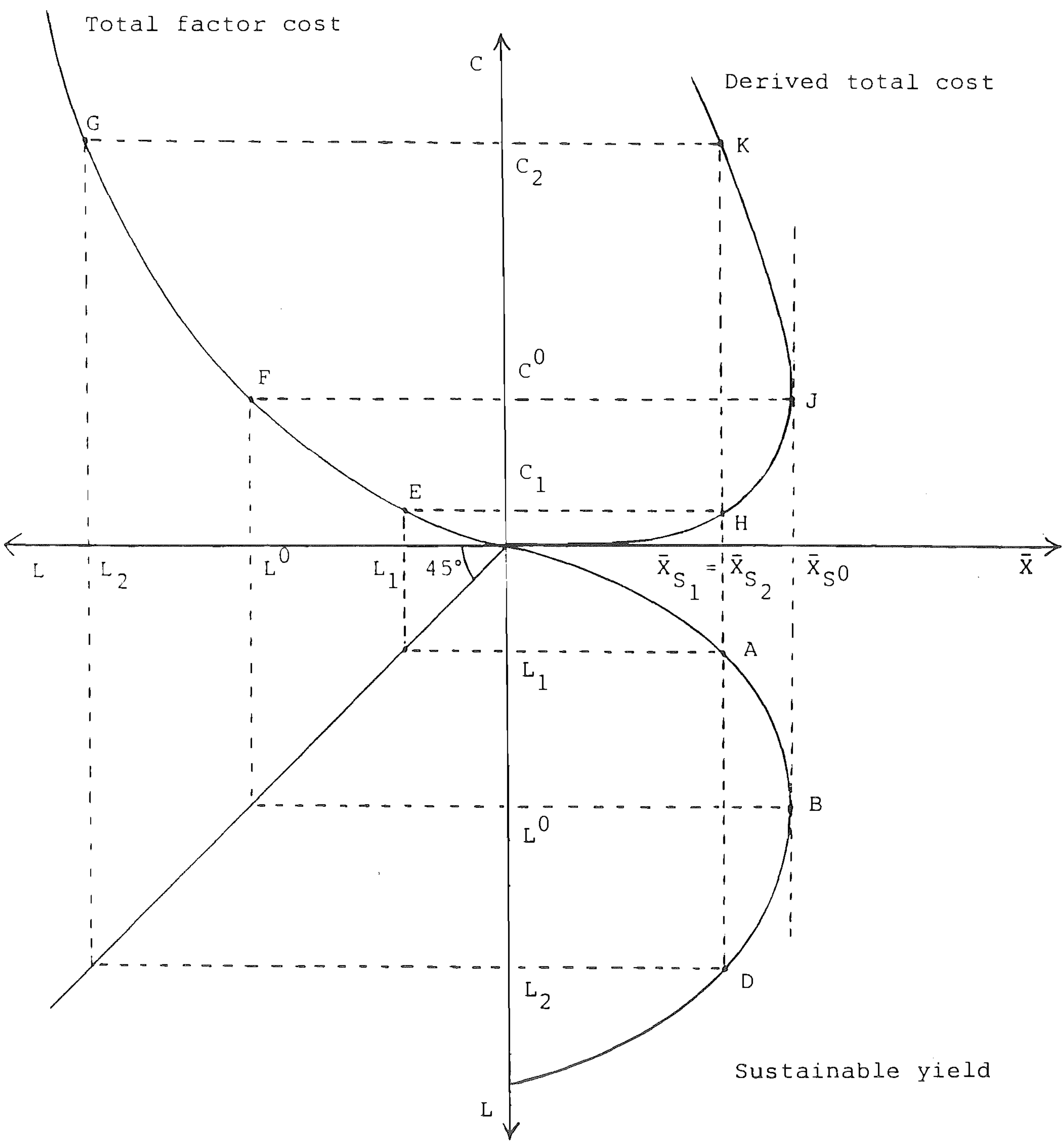
Curve 0CEH shows sustainable revenue $\overline{R}_S$ as a function of labour input to the fishery. The open-access competitive equilibrium occurs at E with a total labour input of 0B > 0G where 0G is the technically efficient amount of labour required to harvest the MSY. At E, total revenue and total cost are equal at level 0A and hence no surplus is generated in the fishery.[2]

The open-access equilibrium at E is contrasted with point C. At C, marginal revenue and marginal cost are equated and resource rent is maximized at CD. This is the point that would be chosen by a sole owner of the fishery and can be achieved by restricting labour input to 0F.

A more interesting analysis, which permits a direct comparison of results from output and input restrictions, uses the sustainable yield curve from Figure 4-3-3 in a four quadrant diagram to derive a total cost curve [Copes, 1970; 70]. This is derived in Figure 4-3-5 overleaf.

Quadrant two shows total factor cost in relation to factor inputs. Assuming competitive factor markets, increments to total factor cost represent marginal opportunity costs of the factors in alternative uses. It is variously shown as linear [Anderson, 1977; 81] or, as in this case, a continuously increasing function [Copes, 1970; 70]. The sustainable yield curve from Figure 4-3-3 is shown in quadrant four. Combining these gives the total cost of output in quadrant one. The derivation of this relationship is as follows.

Figure 4-3-5: Sustainable total cost curve

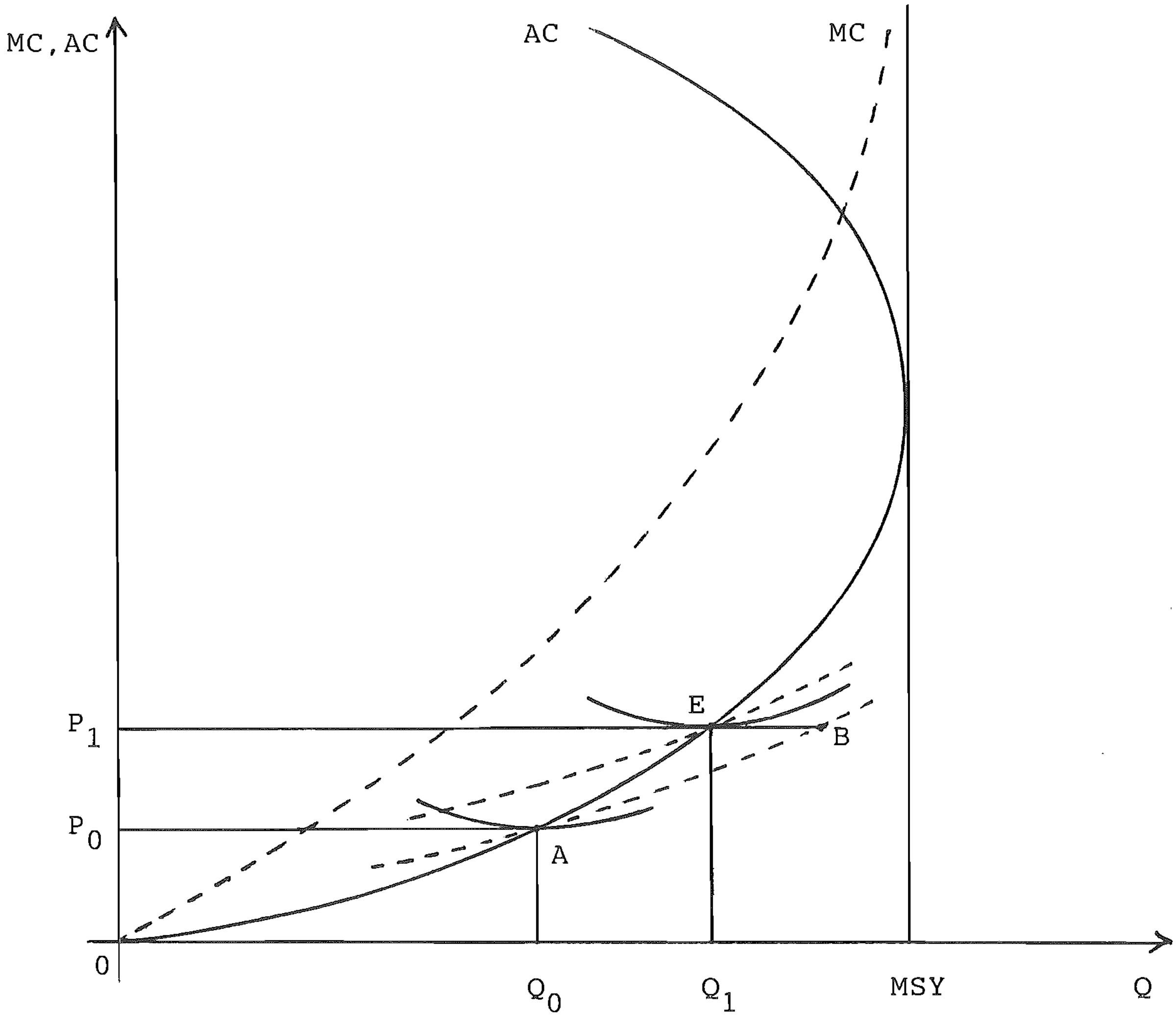



Point A in quadrant four is associated with stock level $S_1$. From Figure 4-3-3 this gives a sustainable harvest of $\bar{X}_{S_1}$ which requires $L_1$ units of labour to produce. The total cost of this is shown as $C_1$ by point E on the total factor-cost function which in turn gives point H on the derived total output-cost function in quadrant one. An increase in labour input to F raises total input costs to $C^0$. Here the labour input $L^0$ is consistent with a smaller population size $S^0$ which generates the MSY of $\bar{X}_{S^0}$ giving point J in quadrant one. Further increases in labour inputs act to increase costs but decrease sustainable output leading to a situation such as points G, D and K. From the derived total cost function, marginal and average cost curves can be found. Average cost rises throughout irrespective of whether the total factor cost function is linear or increasing because of the steady-state stock/yield relationship while marginal cost is positive and above it until the MSY output level.

Quadrant one in Figure 4-3-5 shows that each catch level other than the MSY is associated with two distinct levels of total cost and hence with two levels of average cost also. This produces a backward-bending average cost curve which is shown in Figure 4-3-6 overleaf.

Under open-access fishing, competition ensures that factors enter the industry until the marginal opportunity cost of production equals the market price and no rent is yielded from the resource. Long-run competitive equilibrium output then occurs where long-run average cost is equated with the market price. As output price changes, the long-run average cost curve in Figure 4-3-6 traces out the long-run, or steady-state, supply curve for the open-access fishery. The dynamics which generate this long-run supply curve are similar to those underlying Figure 4-3-3

Figure 4-3-6: Steady-state supply curve




and follow from the impact of stock changes on harvesting costs [Dnes, 1985; 163-164].

Beginning from an equilibrium position at point A, the possibility of economic profits resulting from a price rise to $P_1$, induces entry. Of itself, this would move the fishery to point B on the short-run marginal cost curve. This point, however, corresponds to a catch level greater than the sustainable yield and the resultant population effects act to increase average fishing costs. This process continues until a new zero profit steady-state equilibrium is reached at E where the stock size and extraction costs are consistent with the new price level.

The upward sloping dashed line in Figure 4-3-6 represents long-run marginal cost. As shown, this is well-defined for the upward sloping segment of the supply curve however it becomes undefined at the MSY and is negative for the backward sloping portion of the curve.

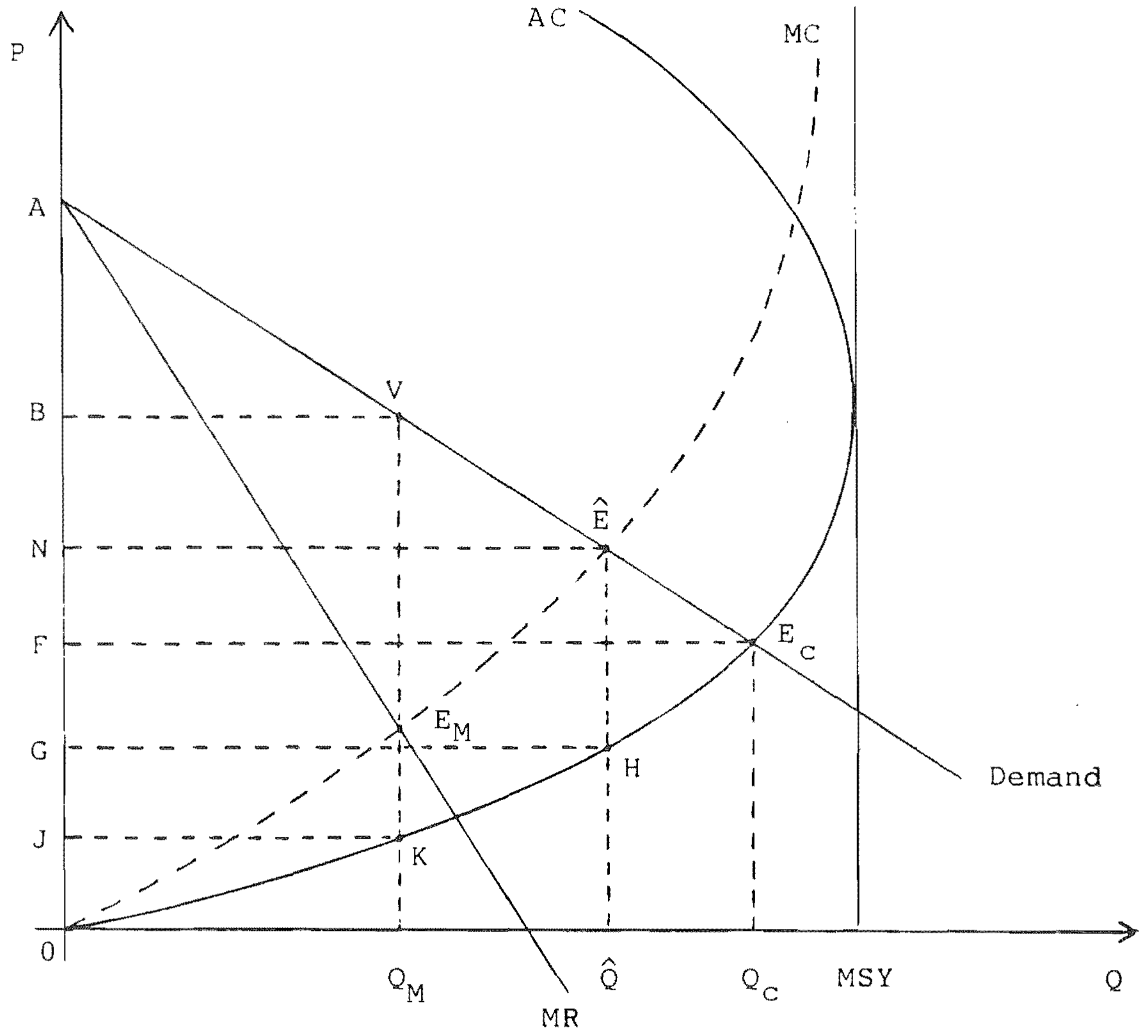
This backward sloping nature of the supply curve raises the possibility of multiple equilibria if confronted by a negatively sloped linear demand curve. The stability and dynamics of these various equilibria under conditions of static and changing demand are analysed in Copes [1970; 72-75]. For the purposes of this analysis, it is assumed that conditions are such as to exclude multiple equilibria.

Within the context of this model Copes [1972; 146-149] derives socially optimal levels of output for downward sloping and perfectly elastic demand curves. Depending on the initial position, these may suggest an expansion of output from the open-access equilibrium level, but they necessarily involve a reduction in factor input to the fishery.

Superimposing a downward sloping demand or average revenue function and its corresponding marginal curve on the long-run average and marginal cost curves from Figure 4-3-6 allows this derivation. This is illustrated in Figure 4-3-7 overleaf.

Initially it is assumed that the intersection of supply and demand occurs on the upward sloping segment of the supply curve. This point $E_C$ shows the open-access competitive equilibrium where price equals average cost and the resource yields zero rent, market revenue being fully absorbed by production costs.

Figure 4-3-7: Competitive equilibrium in an open-access fishery




The consumer surplus of $AE_CF$ provides a net social benefit from the fishery but the output level of $Q_C$ at point $E_C$ represents an overinvestment in fishing activity in terms of the maximization of social welfare. The demand curve represents the marginal social revenue from fishery output and, in the absence of any disparity between private and social opportunity costs of labour, the marginal cost curve also shows marginal social costs.

The socially optimal output from the fishery is then $\hat{Q}$ which corresponds to $\hat{E}$ where marginal social revenue and marginal social cost are equated. This involves a reduction in output of $Q_C\hat{Q}$ units and an increase in market price which creates a divergence between market revenue and production cost. This amount of NÊHG accrues to the resource in the form of rent and may be appropriated in a number of ways. The change in market price reduces consumer surplus by NÊ$E_C$F but this is outweighed by the increase in rent, the social optimum being the quantity where the sum of the two is maximized.

The remaining point $E_M$ represents the optimal output level for a single resource owner or producers' monopoly. At this point, marginal revenue equals marginal cost and resource rent is maximized at amount BVKJ.

Copes further modifies this model to allow for differences in opportunity costs between intramarginal factor inputs. This creates a distinction between private and social costs. "Social" average and marginal cost curves are constructed by subtracting out the producer surplus from the market curves. This separates the objectives of producers and resource owners, and enables the derivation of optimal output levels for differing interest groups [Copes, 1972; 154-159]. Assuming no such divergences, Figure 4-3-7 shows the equilibrium and/or optimal output levels which correspond to various market structures as derived above.

## 4-4 THE EFFECTS OF REGULATION ON AN OPEN-ACCESS FISHERY

Beginning from the competitive open-access equilibrium at $E_C$ in Figure 4-3-7, in a world of costless regulation and complete compliance, either of the remaining two points, $\hat{E}$ and $E_M$, can be attained by an appropriate regime of taxes and/or quotas, the exact mix depending on the incentives of the agents and regulator involved [Anderson, 1977; Clark, 1982; Scott and Southey, 1970]. The informational requirements of a tax and a quota are essentially the same in all cases [Mirman and Spulber, 1985; 732, and Weitzman, 1974; 478] and a tax and tradeable output quota which both achieve any given regulated equilibrium output level are identical in terms of economic efficiency [Clark, 1982; 283].

Regulation in practice however is neither costless nor merits perfect compliance. Compliance occurs only as a result of a rational decision concerning the private costs and benefits of so doing. Without penalties to deter non-compliance, and excluding the case of monopoly ownership of the resource, individual profit-maximizing behaviour will result in the open-access competitive equilibrium harvest irrespective of any regulatory measures introduced. The regulatory agency in deciding what controls, if any, to impose, must offset the costs incurred in the enforcement process against the benefits which are apparent in Figure 4-3-7.

The regulator is assumed to control both inputs to and outputs from the fishery by means of tradeable output quotas. Quotas may be issued at the industry level and auctioned to individual operators or given directly to individual firms on the basis of historical outputs and then traded if required. Indeed, all initial arbitrary quota assignments are equivalent on efficiency grounds given tradeability and perfect capital markets. Their implications for income distribution, however, are not, and this can be an important factor in determining the method to be used.

The regulatory model used here is essentially the same as that introduced in Chapter Two. It is assumed, at the aggregate level, that the regulatory agency imposes a quota of R units which is less than the competitive open-access output level and remains binding on industry decisions. The quota is enforced by means of a constant unit rate fine levied on production in excess of individuals' quota holdings. Not all violations are necessarily detected however. The probability of detection and prosecution ($\rho$) is assumed to be a function of the labour devoted to enforcement, similar in form to production functions elsewhere in the economy. Thus

$$(4\text{-}4\text{-}1) \quad \rho = \rho(L_e);\ \rho' > 0,\ \rho'' < 0;\ \rho(0) = 0,\ \rho(\bar{L}_e) = 1$$

where $L_e$ is the amount of labour devoted to enforcement, and

$$(4\text{-}4\text{-}2) \quad 0 \leq L_e \leq \bar{L}_e < L$$

where $\bar{L}_e$ is the minimum level of enforcement which ensures a probability of detection of unity. It is assumed that it is not necessary to devote all of the economy's resources to enforcement activities in order to ensure that all violations are punished. The expected penalty payable at any level of illegal production (Q-R) is then a function of the unit rate fine (f) and the amount of labour devoted to enforcement.

$$(4\text{-}4\text{-}3) \quad EP(Q,R,f,L_e) = f \cdot \rho(L_e)[Q-R], Q \geq R; \; EP(Q,R,0,L_e)$$
$$= EP(Q,R,f,0) = EP(k,k,f,L_e) = 0, \quad 0 < k < Q_c$$

Following the analysis of Chapters Two and Three, it is assumed that the fine is given in the regulator's enabling statute. The regulator then controls the expected penalty by varying the size of its enforcement effort.

The expected penalty shown in (4-4-3) is some constant unit-rate amount which, given the value of the unit-rate fine, is uniquely determined by the level of enforcement activity. Given this expected penalty, rational profit-maximizing agents consider the expected costs and benefits of exceeding their quota and a regulated equilibrium occurs when the supply and demand decisions of individual agents within the regulatory environment are equated. The regulated equilibrim output level (Q) is therefore a function of the size of the quota and the level of enforcement activity. This is defined for fixed values of some as yet unspecified parameters contained in the parameter vector Z. This vector contains factors such as the level of consumer income which affect the demand and/ or supply curves for the regulated product and hence alter the regulated equilibrium at any value of the policy variables. Thus

$$(4\text{-}4\text{-}4) \qquad Q = Q(L_e, R, Z)$$

The effect of introducing regulatory controls into the fishery can be determined from an examination of the properties of (4-4-4).

Following Section 2-6, the expected penalty becomes an additional cost of production for any output produced by the industry regardless of whether or not individual fishermen exceed their quotas. If a quota is exceeded, the illegal output incurs the expected unit fine in addition to the normal technical costs of production. As the quota is tradeable, production cost for all infra-quota units is similarly affected. This increased unit cost reflects the minimum amount payable for illegal production or the maximum amount paid to acquire the legal right to produce and consequently represents the opportunity cost of any level of production in the regulated environment.

Using the argument of the previous paragraph, the average cost of production at every level of output rises by the amount of the per-unit expected penalty $\lambda$. From (4-4-3)

$$\lambda(f,L_e) = f.\rho(L_e) \qquad (4\text{-}4\text{-}5)$$

Recalling the free entry or open-access assumption, competitive equilibrium occurs when price equals average production costs and the industry earns zero profit. In the regulated environment therefore

$$P = \frac{C(Q)}{Q} + \lambda(f,L_e) \qquad (4\text{-}4\text{-}6)$$

where P is the supply price at any output given the size of the unit rate expected penalty and C(Q) is the cost function derived from the interaction between labour input and stock size which generates the average cost curve in Figure 4-3-7.

(1) Regulation of a "Low-Cost" Fishery

PROPOSITION 4-4-1: Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a strictly binding and tradeable output quota enforced by means of a constant flat-rate per unit expected monetary penalty. Asssuming that the open-access equilibrium in the unregulated environment occurs in the positively sloped region of the industry average cost curve:

(i) An increase in enforcement, such that the expected penalty remains binding on behaviour,[3] increases market equilibrium price and reduces the regulated equilibrium output level of the fishery.

(ii) A change in the size of available quota, such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, has no effect on the regulated equilibrium of the fishery.

Proof.

(i) The regulated equilibrium output level of the fishery in (4-4-4) emerges at the price which equates the industry's penalty-inclusive supply with market demand. Thus at the regulated equilibrium

(4-4-7) $$Q^S(P(f,L_e),L_e,R) - Q^D(P(f,L_e)) = 0$$

where $Q^S$ represents the industry supply curve which corresponds to the AC curve in Figure 4-3-7 and $Q^D$ represents the market demand curve faced by the industry.[4]

Totally differentiating (4-4-7) gives

(4-4-8) $$\frac{\partial Q^S}{\partial P}\frac{\partial P}{\partial L_e} + \frac{\partial Q^S}{\partial L_e} + \frac{\partial Q^S}{\partial R} - \frac{\partial Q^D}{\partial P}\frac{\partial P}{\partial L_e} = 0$$

Holding the quota constant and rearranging (4-4-8) reveals that

(4-4-9) $$\frac{\partial P}{\partial L_e} = -\frac{\partial Q^S/\partial L_e}{\frac{\partial Q^S}{\partial P} - \frac{\partial Q^D}{\partial P}}$$

From the right-hand side of (4-4-6), an increase in the level of enforcement activity raises the penalty-inclusive average cost of production at any output level. From individual profit-maximizing behaviour, the supply price at any output level rises by the amount of increase in the expected penalty rate or alternatively supply falls at every market price for which it is strictly positive and hence $\partial Q^S/\partial L_e < 0$. Following the assumptions of the Proposition, $\partial Q^S/\partial P > 0$ and $\partial Q^D/\partial P < 0$. Using these results in (4-4-9) reveals that $\partial P/\partial L_e > 0$ and therefore an increase in enforcement, such that the expected penalty remains binding on behaviour, increases market equilibrium price.

The regulated equilibrium of the fishery lies somewhere on the market demand curve. Multiplying (4-4-9) by the price responsiveness of demand then, given the assumption that the demand curve is negatively sloped, shows that the increase in enforcement reduces the regulated equilibrium output of the fishery.

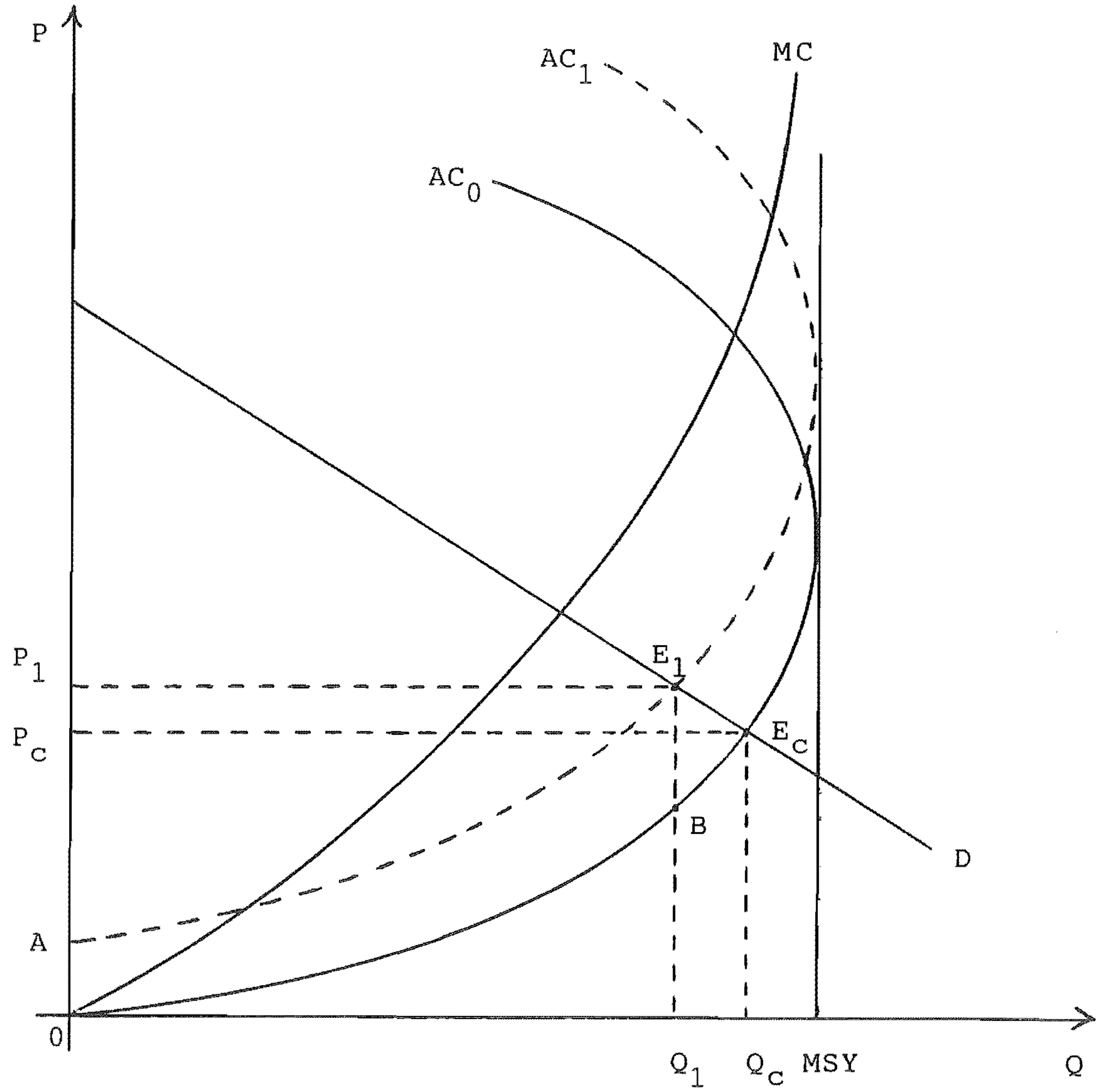
(ii) From (4-4-6) the industry supply price is independent of any change in the size of the output quota provided that the quota does not exceed the regulated equilibrium output level of the fishery generated by the level of enforcement activity. Following on therefore, from (4-4-8), holding the level of enforcement constant, $\partial Q^S/\partial R = 0$ and neither the market price nor the regulated equilibrium output level of the fishery are affected by a change in available quota such that the expected penalty remains binding on behaviour.

□

Result (i) of Proposition 4-4-1 demonstrates the deterrent effect of enforcement within the fishery and corresponds to the similar effect derived in Proposition 2-6-2 in the case of the regulation of a negative-externality generating industry. The result here is illustrated in Figure 4-4-1 overleaf.

Beginning from the unregulated open-access equilibrium at $E_c$, a policy consisting of some quota level $R < Q_c$ and an expected penalty $\lambda$, which corresponds to some level of enforcement and given unit-rate fine and which is binding on behaviour in conjunction with the quota level R, is introduced. The expected penalty increases the average cost of production at every quantity shifting the supply

Figure 4-4-1: The effect of regulation by output quota on a "low-cost", open-access fishery.




curve vertically by the amount of the per unit expected penalty $\lambda$. The intersection of the demand curve and the penalty-inclusive supply curve gives a regulated equilibrium at $E_1$. Market equilibrium price rises to $P_1$ and quantity falls to $Q_1$, the precise effects being dependent on the price responsiveness of the curves.

Any change in the quota level R, such that $R < Q_1$, does not affect the regulated equilibrium of the fishery because it leaves the marginal expected penalty unchanged. This is analogous to result (ii) Proposition 2-7-2. Recalling that all units of output incur the additional cost of the expected penalty, the area $0AE_1B$, which corresponds to AKJT in Figure 2-7-1, shows the aggregate value of expected penalty at the regulated equilibrium output level. The size of the quota then determines the distribution of this amount between rents which accrue to quota holders and fines paid for illegal production.

(2) <u>Regulation of a "High-Cost" Fishery</u>

As shown in the derivation of Figure 4-3-6, there are two methods of producing any steady-state catch other than the MSY which differ in the amount of factor input and hence in the level of total production cost. Figure 4-4-1 illustrates the effect of an output quota on a "low-cost" open-access fishery. It is possible, however, that demand and cost conditions are such as to produce an unregulated open-access equilibrium in the backward-bending portion of the supply curve. This case is described here as a "high-cost" fishery.

<u>PROPOSITION 4-4-2:</u> <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a strictly binding and tradeable output quota enforced by means of a constant flat-rate per unit expected monetary penalty. Assuming that the open-access</u>

equilibrium in the unregulated environment occurs in the negatively sloped region of the industry average cost curve and that the slope of the backward-bending portion of the industry supply curve is more negative than that of the demand curve:

(i) An increase in enforcement, such that the expected penalty remains binding on behaviour,[5] reduces market equilibrium price and increases the regulated equilibrium output level of the fishery. This continues until the level of enforcement which generates a regulated equilibrium at the MSY is attained.

(ii) A change in the amount of available quota, such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, has no effect on the regulated equilibrium of the fishery.

Proof.

(i) The assumption concerning the relative slopes of the demand curve and the backward-bending portion of the industry average cost curve, which from the open-access assumption is the long-run supply curve of the fishery, is sufficient to ensure that any open-access equilibrium which occurs on the backward-bending portion of the supply curve is unique.

From (4-4-9) the effect of a change in enforcement on the market equilibrium price is

$$\text{(4-4-10)} \qquad \frac{\partial P}{\partial L_e} = - \frac{\partial Q^S/\partial L_e}{\frac{\partial Q^S}{\partial P} - \frac{\partial Q^D}{\partial P}}$$

The assumption of a negatively sloped market demand curve ensures that $\partial Q^D/\partial P < 0$ as, by the assumption that open-access equilibrium occurs in the backward-bending portion of the supply curve, is $\partial Q^S/\partial P$ also. The sign of the denominator of the right-hand side of (4-4-10) then is not unambiguous. From the assumption contained in the Proposition concerning the relative slopes of the supply and demand curves however,

$$\text{(4-4-11)} \qquad 0 > \frac{\partial Q^S}{\partial P} > \frac{\partial Q^D}{\partial P}$$

and hence the denominator is positive.

From (4-4-6), using (4-4-1) and (4-4-5), an increase in enforcement increases the size of the unit rate expected penalty and thus raises the supply price at any output level. Given that the supply curve is negatively sloped in the neighbourhood of the unregulated open-access equilibrium, this is equivalent to a rightward shift in the supply curve at the given market price and hence $\partial Q^S/\partial L_e > 0$. Using these results in (4-4-10) reveals that $\partial P/\partial L_e < 0$ and an increase in enforcement, such that the expected penalty remains binding on behaviour, reduces the market equilibrium price in the fishery. Multiplying (4-4-10) by the price responsiveness of demand then shows the consequent rise in the regulated equilibrium output level which occurs as a result of the increase in enforcement.

Given that, by assumption in the model, the quota does not exceed the unregulated open-access equilibrium output level of the fishery, the expected penalty remains binding on behaviour for all levels of enforcement such that the regulated equilibrium occurs on the backward-bending portion of the industry supply curve. This is the case until the MSY is attained. Hence, from (4-4-10), increases in enforcement reduce market equilibrium price and raise regulated equilibrium output until the level of enforcement which generates a regulated equilibrium at the MSY is reached.

(ii) The quota, being constrained to be no greater than the unregulated open-access output level, is strictly less than any regulated equilibrium that occurs on the backward portion of the supply curve. Given therefore that the marginal expected penalty is constant and independent of the quota size, it follows from (4-4-6) and (4-4-8) that neither the market equilibrium price nor the regulated equilibrium output level are affected by a change in the amount of available quota.

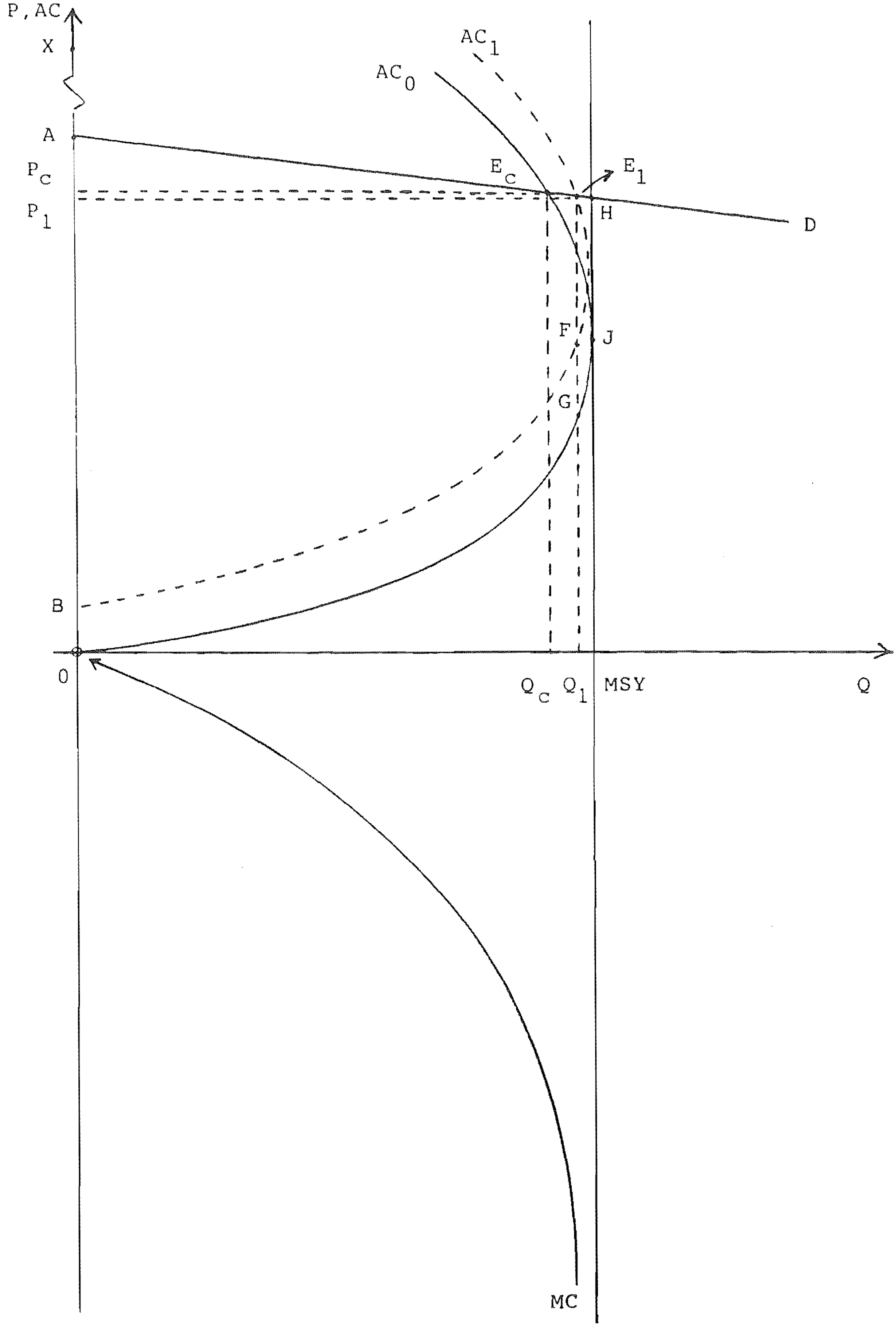
□

Result (i) of Proposition 4-4-2 is contrary to the deterrent effect of enforcement. The analysis of Section 2-7 revealed that this type of result could occur with forms of the expected penalty which violated the concavity of firms' profit functions. Here the result emerges from the dynamics of the stock/yield relationship. Beginning from a "high cost" open-access equilibrium, which corresponds to a point such as B in Figure 4-3-3 or D in Figure 4-3-5, the increase in penalty-inclusive production cost reduces the labour input to the fishery and hence reduces the short-run harvest. This allows the stock level to build up until a new regulated equilibrium is reached at a higher output level. Figure 4-4-2 illustrates the situation.

For the backward-bending portion of the supply curve, the corresponding marginal cost curve is negative apart from the vertical axis intercept at X which shows the cost of producing the initial (or final) unit of output and is an extremely large positive amount.[6] As in the case of the low-cost fishery shown in Figure 4-4-1, marginal cost becomes undefined at the maximum sustainable yield.

Beginning from the unregulated open-access equilibrium at $E_c$, the introduction of an output quota R, such that $R < Q_c$, and some level of enforcement $L_e$, leads to a vertical shift in the industry supply curve from $AC_0$ to $AC_1$. In the region where the supply curve is negatively

Figure 4-4-2: The effect of regulation by output quota on a "high-cost" open-access fishery.



sloped this leads, in effect, to a rightward shift in the regulated supply curve at every price level. A new regulated equilibrium is established at $E_1$. Market price has fallen from $P_c$ to $P_1$ and the output from the fishery has increased from $Q_c$ to $Q_1$. Extending the argument, there is some expected penalty, shown on Figure 4-4-2 as the unit rate HJ, which generates a regulated equilibrium at the maximum sustainable yield output level. This regulated equilibrium occurs at point H in Figure 4-4-2.

Recalling again that all output incurs the unit rate expected penalty, the area 0BFG, which corresponds to $0AE_1B$ in Figure 4-4-1, shows the aggregate size of the expected penalty at the regulated equilibrium output level. Any change in the amount of available quota serves to redistribute this amount between rents to quota holders and fines paid for illegal production. Given the assumption that the quota does not exceed $Q_c$, fine payments at any regulated equilibrium output level such as $Q_1$ will be strictly positive.

Proposition 4-4-2 and Figure 4-4-2 apply to the regulation, by enforced output quota, of a "high-cost" open-access fishery. Further increases in enforcement, beyond the level which generates a regulated equilibrium at the MSY, produce regulated equilibria on the upward-sloping portion of the penalty-inclusive industry supply curve. Here the analysis of Proposition 4-4-1 and Figure 4-4-1 applies.

Section 4-4 has derived the concept of a regulated equilibrium in an open-access fishery subject to an output quota enforced by means of a constant unit-rate expected monetary penalty and examined the effect of changes in regulatory policy on this equilibrium. In the following sections the analysis examines the characteristics of optimal regulatory policy under differing assumptions concerning the objectives of the regulator.

Following Chapter Three, two regulatory hypotheses are treated; the Naive Public Interest (NPIT) hypothesis and the Capture Theory (CT) approach. It is assumed that the NPIT regulator seeks to maximize social welfare whereas the CT regulator acts so as to maximize industry profits. In a world of costless regulation and perfect compliance a NPIT regulator would operate the fishery at $\hat{E}$ in Figure 4-3-7 while a CT regulator would restrict output to the pure monopoly level $Q_M$. As earlier stated, however, regulation is not costless and neither is compliance complete. The regulator must therefore compare costs and benefits in determining its optimal regulatory policy.

## 4-5 THE OBJECTIVES OF THE REGULATOR: NPIT REGULATION

The necessity for enforcement of a regulation ensures that there is a resource cost ($\omega$) associated with controlling the output of the fishery. Following 3-2-2, this is given as

$$(4\text{-}5\text{-}1) \qquad \omega = \omega(L_e) \ ; \ \omega(0) = 0 \ , \ \omega'(L_e) > 0, \ \omega''(L_e) \geq 0$$

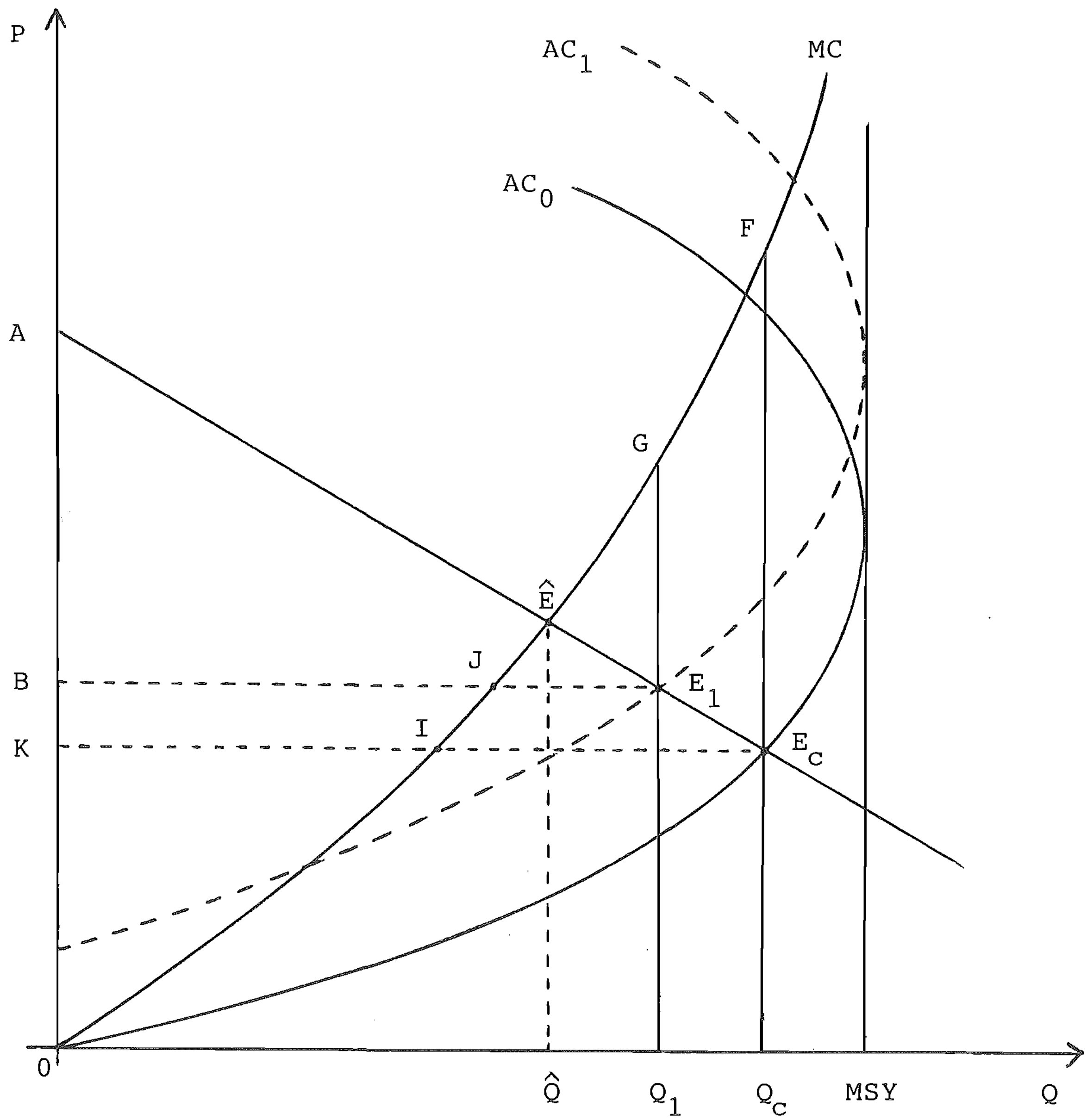
The marginal resource cost is positive and non-decreasing in the level of enforcement activity. It is assumed that this resource cost is funded by a poll tax of some form, the contributors to which are as yet unspecified. Any fine payments incurred by the industry, or rents which accrue to quota holders, represent transfer payments within the economy and, as such, are irrelevant to the decisions of the NPIT regulator. The marginal cost curve MC in Figure 4-3-7 then represents social marginal cost in a regulated "low-cost" fishery and is used to determine the NPIT-optimal regulatory policy.

The analysis of the costs and benefits of regulation is based on an examination of its effects on consumer and producer surplus as defined in the Marshallian tradition. This is illustrated in Figure 4-5-1 in the case of NPIT regulation of a "low-cost" fishery.

Beginning from the unregulated open-access equilibrium point $E_c$, a policy consisting of a quota $R < Q_c$ and an expected per unit penalty of $\lambda$ is introduced. This expected penalty increases the average cost of production at every quantity, shifting the supply curve vertically by the amount $\lambda$. The intersection of the demand curve and the penalty-inclusive supply curve produces a regulated equilibrium at a point such as $E_1$, the precise effect on price and quantity being dependent on the price elasticities of the curves.

At $E_c$, consumers enjoy a surplus of amount $AE_cK$ while producers earn profits of 0KI on early units but suffer losses of $IFE_c$ on subsequent units. These amounts of producer surplus exactly offset each other by the open-access assumption of zero profit. The shift to $E_1$ reduces consumer surplus to $AE_1B$ and increases producer surplus to $0BJ - JGE_1$, resulting in a net welfare gain of the shaded area $GE_1E_cF$. This aggregate welfare gain in restricting output from $Q_c$ to $Q_1$ must be compared with the aggregate

Figure 4-5-1: NPIT Regulation of a "low-cost" open-access fishery




enforcement cost necessary to achieve the change in output.

(1) Optimal Policy in a "Low-Cost" Fishery

The aim of NPIT regulation therefore is to recover lost potential aggregate surplus by reducing output from its unregulated open-access level. This is given in

(4-5-2) below.

$$\text{(4-5-2)} \quad \underset{L_e, R}{\text{Maximize}} \quad V_{R_e}(L_e, R, Z) = \int_{Q_c}^{Q(L_e, R, Z)} [h(Q, Y) - C_1(Q, \phi)] dQ - \omega(L_e) \qquad \text{subject to } R \leq Q_c$$

where the restriction $R \leq Q_c$ corresponds to the assumption from Chapters Two and Three that the quota is binding on behaviour.

The integral term in (4-5-2) represents the amount of recovered potential surplus where

$$\text{(4-5-3)} \qquad h(Q,Y) \; ; \; h_1 < 0, \; h_2 > 0$$

is the inverse market demand function, and

$$\text{(4-5-4)} \qquad C_1(Q,\phi) \; ; \; C_1 > 0, \; C_{11} > 0, \; C_{12} > 0$$

is the industry marginal cost curve for a "low-cost" fishery. The parameter Y in (4-5-3) represents the level of aggregate consumer income, the condition $h_2 > 0$ reflecting the assumption that the output of the fishery is a normal good. The parameter $\phi$ in (4-5-4) represents some factor which increases marginal cost at any output level. Finally, the term $\omega(L_e)$ is the resource cost of enforcement as defined in (4-5-1).

Following the assumptions contained in (4-5-1), (4-5-3) and (4-5-4), and the results of Proposition 4-4-1, the NPIT regulator's objective function in (4-5-2) is concave. The first-order conditions of (4-5-2) then provide a maximum. Differentiating (4-5-2) gives

$$(4\text{-}5\text{-}5)\quad \frac{\partial V_{R_e}}{\partial L_e} = [h(Q(L_e^*,R^*,Z),Y) - C_1(Q(L_e^*,R^*,Z),\phi)]\frac{\partial Q}{\partial L_e} - \omega'(L_e^*) \leq 0 \qquad ; < \text{ only if } L_e^* = 0$$

and

$$(4\text{-}5\text{-}6)\quad \frac{\partial V_{R_e}}{\partial R} = [h(Q(L_e^*,R^*,Z),Y) - C_1(Q(L_e^*,R^*,Z),\phi)]\frac{\partial Q}{\partial R} \leq 0 \qquad ; < \text{ only if } R^* = 0$$

The simultaneous solution of (4-5-5) and (4-5-6) gives the NPIT-optimal levels of enforcement and quota, $L_e^*$ and $R^*$ respectively. Using (4-4-4) this gives the NPIT-optimal regulated equilibrium $Q^*$ for fixed values of the parameters contained in the parameter vector $Z$ where

$$(4\text{-}5\text{-}7)\qquad Q^* = Q(L_e^*,R^*,Z)$$

<u>LEMMA 4-5-1</u>: <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of a constant flat rate per unit expected penalty which is binding on behaviour. At the NPIT-optimal regulated equilibrium output level $Q^*$, the demand price is less than marginal cost.</u>

<u>Proof</u>.

The first-order condition (4-5-5) shows that, at any non-zero level of enforcement which is optimal for the NPIT regulator, $\partial V_{R_e}/\partial L_e = 0$. Given the assumption that the marginal expected penalty is binding on behaviour, then, by Proposition 4-4-1, $\partial Q/\partial L_e < 0$. From (4-5-1), the marginal resource cost of any non-zero level of enforcement is positive and therefore $\omega'(L_e^*) > 0$. Using these results on the signs of the component terms of (4-5-5), it is necessary for $\partial V_{R_e} / \partial L_e^* = 0$ that

$$h(Q(L_e^*,R^*,Z),Y) - C_1(Q(L_e^*,R^*,Z),\phi) < 0 \tag{4-5-8}$$

and thus any regulated equilibrium which involves less output than the unregulated open-access level, and which is optimal for the NPIT regulator, occurs at an output level such that marginal production cost exceeds the demand price.

Alternatively, if no enforcement is optimal for the NPIT regulator, the regulated equilibrium will coincide with the unregulated open-access equilibrium. Given the assumption that the open-access output level is socially excessive, condition (4-5-8) holds in this case also. The NPIT-optimal regulated equilibrium therefore always occurs at an output level where marginal production cost exceeds the demand price.

□

<u>PROPOSITION 4-5-1</u>: <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of a constant flat rate per unit expected penalty that is binding on behaviour. Assuming that it is optimal for a NPIT regulator to enforce the quota;</u>

(i) <u>NPIT-optimal enforcement occurs where the marginal social benefit of enforcement is equal to its</u>

marginal social cost. It is not optimal for the NPIT regulator to restrict output to the first-best socially-optimal level.

(ii) Any quota level that does not exceed the regulated equilibrium output level consistent with the NPIT-optimal amount of enforcement, as determined in part (i) of the Proposition, is optimal for the NPIT regulator.

Proof.

(i) The result follows from (4-5-5) and Lemma 4-5-1. Using Lemma 4-5-1, the bracketed term $[h(.)-C_1(.)]$ in (4-5-5) shows the unit loss in aggregate surplus at any output level in the relevant range. Multiplying by the term $\partial Q/\partial L_e$, which by Proposition 4-4-1 is negative, gives the value of potential aggregate surplus that is recovered. This represents the marginal social benefit of enforcement.

The remaining term in (4-5-5) is $\omega'(L_e^*)$. This is the marginal resource cost of enforcement which, following (4-5-1), is positive. Given that other 'costs' associated with enforcement, such as fine payments and quota rentals, are transfer payments within the economy, the marginal resource cost represents the marginal social cost of enforcement.

Given that the first-order condition (4-5-5) requires that $\partial V_{R_e}/\partial L_e = 0$ at any non-zero NPIT-optimal enforcement level, the above discussion shows that the NPIT-optimal enforcement level occurs when the marginal social benefit and marginal social cost of enforcement are equated.

At the first-best socially optimal output level of the fishery, shown by $\hat{Q}$ in Figure 4-5-1, demand price and marginal production cost are equal. From Lemma 4-5-1 this output level is not optimal for a NPIT regulator given the existence of a positive marginal resource cost of enforcement. Given that the demand price falls with output and that the marginal cost of production increases with output, the NPIT-optimal regulated equilibrium output level, by Lemma 4-5-1, occurs at some $Q > \hat{Q}$.

(ii) Following Proposition 4-4-1, with a constant flat rate per unit expected penalty, the marginal expected penalty with respect to output, and hence the regulated equilibrium output level of the fishery, is unaffected by any change in quota such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement. Thus $\partial Q/\partial R = 0$ and, from (4-5-6), any output quota, which does not exceed the regulated equilibrium output level generated by the NPIT-optimal amount of enforcement as determined in part (i) of the Proposition, is optimal for the NPIT regulator.

□

Following Proposition 4-5-1 and Lemma 4-5-1, and assuming that it is optimal for the NPIT regulator to control the industry, the NPIT-optimal regulated equilibrium $Q^*$, using (4-5-7), occurs at some intermediate output level

$$(4\text{-}5\text{-}9) \qquad \hat{Q} < Q(L_e^*, R^*, Z) < Q_c \ ; \ L_e^* > 0, \ 0 \leq R^* \leq Q(L_e^*, R^*, Z)$$

which is generated by a unique optimal amount of enforcement coupled with any quota that is binding on behaviour.

## (2) Policy Responses to Parameter Changes in a "Low-Cost" Fishery

The parameters contained in Z are explained in Appendix 4-1. The effects of parameter changes on the optimal values of these policy instruments, and hence on the NPIT-optimal regulated equilibrium itself, can be examined by totally differentiating the first-order conditions for a maximum. Given the result from Proposition 4-4-1 that the regulated equilibrium is unaffected by changes in the output quota such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, and assuming that the quota level can be costlessly adjusted by the NPIT regulator, the effect of parameter changes on NPIT-optimal regulatory policy can be determined solely by examining the first-order condition for enforcement.

Totally differentiating (4-5-5) gives

$$(4\text{-}5\text{-}10) \qquad A dL_e^* + B dR^* + E dY + F d\phi = 0$$

where

$$A = [h(.) - C_1(.)]\frac{\partial^2 Q}{\partial L_e^2} + [h_1(.)\frac{\partial Q}{\partial L_e} - C_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial L_e} - \omega''(L_e)$$

$$B = [h(.)-C_1(.)]\frac{\partial^2 Q}{\partial L_e \partial R} + [h_1(.)\frac{\partial Q}{\partial R} - C_{11}(.)\frac{\partial Q}{\partial R}]\frac{\partial Q}{\partial L_e}$$

$$E = [h_1(.)\frac{\partial Q}{\partial Y}+h_2(.)-C_{11}(.)\frac{\partial Q}{\partial Y}]\frac{\partial Q}{\partial L_e} + [h(.)-C_1(.)]\frac{\partial^2 Q}{\partial L_e \partial Y}$$

$$F = [h_1(.)\frac{\partial Q}{\partial \phi} - C_{11}(.)\frac{\partial Q}{\partial \phi} - C_{12}(.)]\frac{\partial Q}{\partial L_e} + [h(.)-C_1(.)]\frac{\partial^2 Q}{\partial L_e \partial \phi}$$

The first parameter change to be considered concerns the level of aggregate consumer income.

(a) The level of aggregate consumer income.

PROPOSITION 4-5-2: Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of a constant flat rate per unit expected penalty that is binding on behaviour. Suppose also that, throughout the parameter-induced changes in optimal enforcement, the quota remains such that it does not exceed the regulated equilibrium output level consistent with the amount of enforcement, and that the NPIT regulator's objective function $V_{R_e}$ is concave in enforcement. Assuming that a change in aggregate consumer income leaves the marginal productivity of the regulator's policy instruments unchanged, an increase in income which increases demand for the output of the fishery reduces (increases) the NPIT-optimal level of enforcement if and only if the income-induced increase in demand price at the initial regulated equilibrium exceeds (is less than) the income-induced change in the per unit loss of aggregate surplus on the marginal unit of output.

Proof.

Given the assumption that $V_{R_e}$ is concave in enforcement it follows that term A in (4-5-10) is negative. The assumption that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement throughout the parameter-induced changes in enforcement implies that $\partial Q/\partial R = 0$ and $\partial^2 Q/\partial L_e \partial R = 0$. Therefore the term B in (4-5-10) is zero.

Using (4-5-10) with $d\phi = 0$ gives

$$(4\text{-}5\text{-}11) \qquad \frac{dL_e^*}{dY} = -\frac{E}{A} \gtrless 0 \quad \text{if and only if} \quad E \gtrless 0$$

given the result that $A < 0$.

Following Proposition 4-4-1, $\partial Q/\partial L_e < 0$ while, from (A4-1-4), $\partial Q(.)/\partial Y > 0$ and the increase in aggregate income increases the regulated equilibrium of the fishery. The assumption that the change in income does not affect the marginal productivity of enforcement means that $\partial^2 Q/\partial L_e \partial Y = 0$. Using this information in (4-5-10) reveals that

$$(4\text{-}5\text{-}12) \qquad E \gtrless 0 \quad \text{if and only if} \quad h_2(.) \lessgtr -[h_1(.) - C_{11}(.)]\frac{\partial Q}{\partial Y}$$

The term $h_2(.)$ denotes, following (4-5-3), the income-induced change in the demand price at the initial regulated equilibrium. The term $[h_1(.)-C_{11}(.)]$ shows the unit rate of change in the difference between the demand price $h(.)$ and marginal production cost $C_1(.)$. Given that, from Lemma 4-5-1, marginal cost exceeds the demand price over the relevant output range, multiplying this square-bracketed term by $\partial Q/\partial Y$ gives the income-induced change in the unit loss of social surplus on the marginal unit of output. Applying condition (4-5-12) in (4-5-11), the result of the Proposition clearly follows.

□

The result of this Proposition is the same as that of part (i) of Proposition 3-3-1 as is its interpretation. Multiplying the terms in (4-5-12) by $\partial Q/\partial L_e$ reveals again that the response in NPIT-optimal enforcement to the change in income depends on the income-induced changes in the marginal benefit and marginal cost of enforcement to the regulator. Here, as in Proposition 3-3-1, the result can be complicated by incorporating income-induced changes in the marginal productivity of enforcement.[7]

The second parameter change considered concerns the structure of marginal production cost in the fishery. This is proxied by the parameter $\phi$ which may represent the state of available technology or the mortality rate of the fishery.

(b) The structure of marginal production cost.

PROPOSITION 4-5-3: Under the assumptions in the stem of Proposition 4-5-2, a change in a

<u>parameter which leads to an increased marginal cost of production at any output level, but which does not affect the marginal productivity of enforcement, reduces (increases) the NPIT-optimal level of enforcement if the parameter-induced change in the marginal cost of production at the initial regulated equilibrium is less than (exceeds) the parameter-induced change in the per unit loss of aggregate surplus on the marginal unit of output.</u>

<u>Proof</u>.

Using (4-5-10) with dY = 0 gives

$$\text{(4-5-13)} \qquad \frac{dL_e^*}{d\phi} = -\frac{F}{A} \gtreqless 0 \quad \text{if and only if} \quad F \gtreqless 0$$

given the results on the component terms of (4-5-10) expressed in the proof of Proposition 4-5-2.

Following Proposition 4-4-1, $\partial Q/\partial L_e < 0$ while, from Appendix 3-1, $\partial Q/\partial\phi < 0$ also. The assumption that the change in $\phi$ does not affect the marginal productivity of enforcement means that $\partial^2 Q/\partial L_e \partial\phi = 0$. Using this information in (4-5-10) shows that

$$\text{(4-5-14)} \quad F \gtreqless 0 \quad \text{if and only if} \quad C_{12}(.) \gtreqless [h_1(.)-C_{11}(.)]\frac{\partial Q}{\partial\phi}$$

The term $C_{12}(.)$ denotes, following (4-5-4), the parameter-induced change in marginal production cost at the initial regulated equilibrium. The term $[h_1(.)-C_{11}(.)]$ has the interpretation as for the proof of Proposition 4-5-2.

Multiplying the square-bracketed term by $\partial Q/\partial\phi$ gives the parameter-induced change in the unit loss of aggregate surplus on the marginal unit of output. Applying condition (4-5-14) in (4-5-13), the result of the Proposition clearly follows.

□

The argument concerning the parameter $\phi$ is exactly the converse of that for the change in income. An increase in $\phi$ which raises marginal cost reduces the first-best socially-optimal output level for the fishery. This of itself leads to an increase in the marginal benefit of enforcement at any output level which exceeds $\hat{Q}$ and thus allows NPIT-optimal enforcement activity to be expanded. However, the increased cost structure also leads to a reduction in the regulated equilibrium output level at any given level of enforcement. This reduces the loss of aggregate surplus on the marginal unit of output and thus acts to reduce the marginal benefit of enforcement. As in all previous cases, these two factors must be weighed against each other in order to determine the NPIT-optimal policy response. Parameter-induced changes in the marginal productivity of enforcement may reinforce or reverse the results of the Proposition depending on their sign and relative magnitude.

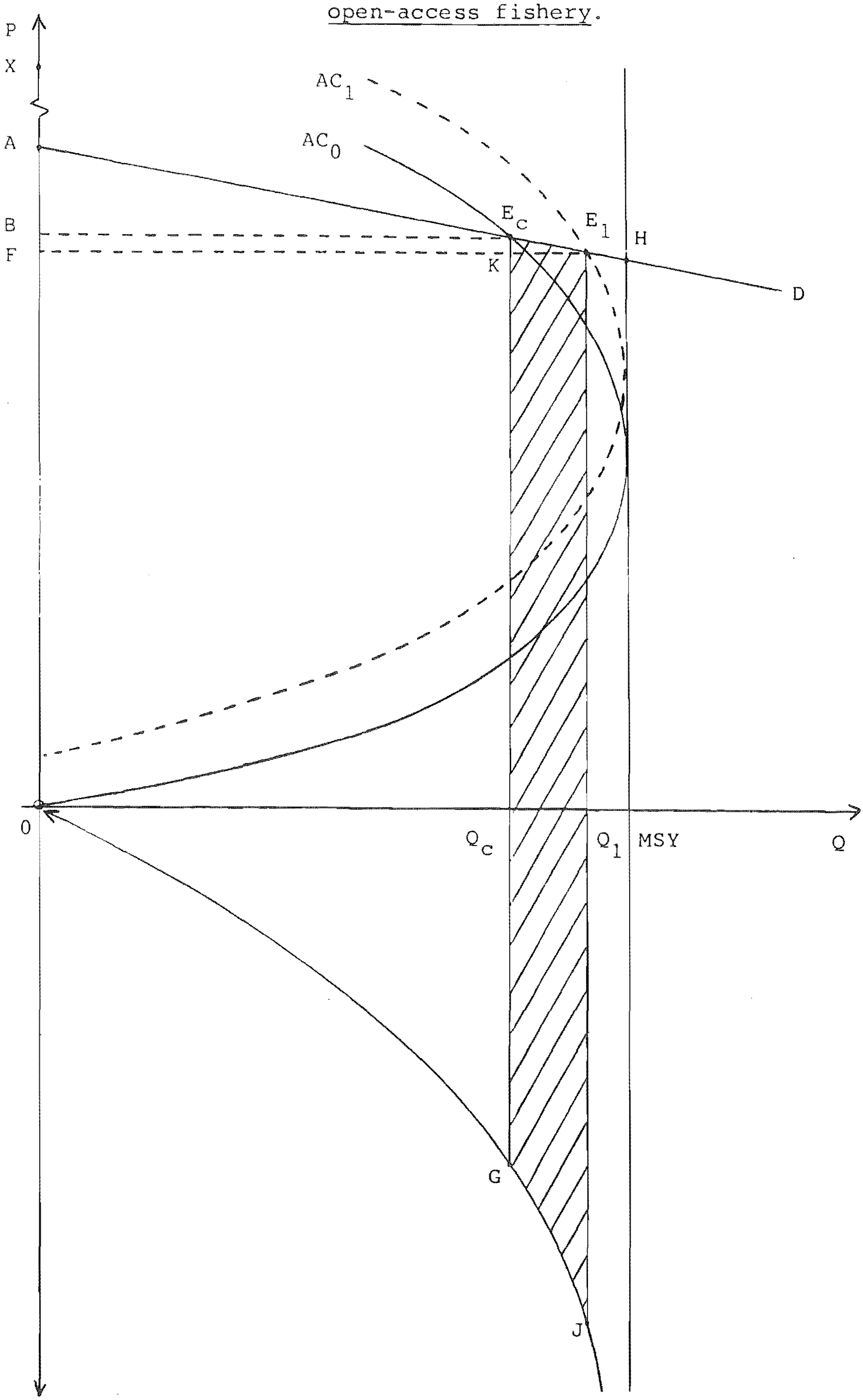
### (3) Optimal Policy in a "High-Cost" Fishery

These results concerning NPIT-optimal regulation of an open-access fishery pertain to the situation of a "low-cost" fishery. In the situation of a "high-cost" fishery, illustrated in Figure 4-4-2, the NPIT regulator's objective is exactly the same as that in the "low-cost" case; namely to maximize aggregate surplus. Thus (4-5-2) remains the expression of the NPIT regulator's objective function. Following Proposition 4-4-2 and Figure 4-4-2, however, the demand price exceeds marginal cost which is negative and the effect of enforcement is to increase the output of the fishery. In this case the quota acts to reduce factor input into the fishery which, given the dynamics in Figure 4-3-3, results in a greater harvest at lower average cost.[8] This is illustrated in Figure 4-5-2 overleaf.

At the initial competitive equilibrium output $Q_c$, in Figure 4-5-2, consumer surplus is $AE_cB$ while producer surplus is $0BE_cG$ less the initial loss of XB where X is the vertical axis intercept of the marginal cost curve. This amount XB exactly offsets the area $0BE_cG$ by the zero profit assumption. Some quota $R < Q_c$ is imposed. Enforcement of the quota shifts the average cost curve vertically by the amount of the marginal expected penalty producing a regulated equilibrium at $E_1$ with output of $Q_1$. Consumer surplus increases to amount $AE_1F$ while producers lose $BE_cKF$ and gain $KE_1JG$. The net result of these changes is an aggregate gain in welfare of the shaded area $GE_cE_1J$.

In this case, the NPIT regulator's objective function, (4-5-2), is no longer necessarily concave in enforcement.[9] The first term of the first-order condition (4-5-5) again shows the marginal benefit of enforcement. As shown in Figure 4-5-2, marginal cost is negative in the region for which the supply curve is negatively sloped. At regulated equilibria $Q_c < Q(L_e,R) < MSY$ which occur on the

Figure 4-5-2: NPIT regulation of a "high-cost" open-access fishery.



backward-bending portion of the industry average cost curve

$$(4\text{-}5\text{-}15) \qquad h(Q(L_e,R),Y) - C_1(Q(L_e,R),\phi) > 0$$

This gives the per unit amount of potential aggregate surplus. From Proposition 4-4-2 $\partial Q/\partial L_e > 0$. Multiplying the left-hand side of (4-5-15) by $\partial Q/\partial L_e$ then gives the marginal benefit of enforcement. From (4-5-8), the derivative of the marginal benefit of enforcement with respect to the level of enforcement is

$$(4\text{-}5\text{-}16) \quad [h(.)-C_1(.)]\frac{\partial^2 Q}{\partial L_e^2} + [h_1(.)\frac{\partial Q}{\partial L_e} - C_{11}(.)\frac{\partial Q}{\partial L_e}]\frac{\partial Q}{\partial L_e}$$

Depending on the relative magnitudes of these two terms the marginal benefit of enforcement may be increasing or decreasing as the level of enforcement activity rises. At the MSY output level marginal production cost is undefined and $\partial Q/\partial L_e = 0$. The marginal benefit of enforcement then approaches some finite limit value at the level of enforcement which generates a regulated equilibrium at point H in Figure 4-5-2 where output is at the MSY. If this limit value exceeds the marginal resource cost of the associated level of enforcement, further enforcement is optimal for the NPIT regulator. In this case, the regulated equilibrium occurs on the upward-sloping portion of the supply curve. NPIT-optimal regulatory policy then proceeds as for Proposition 4-5-1 in the case of a "low-cost" fishery. If, however, the marginal benefit of enforcement is decreasing, or increasing at a slower rate than the marginal resource cost, it is possible that the NPIT-optimal regulated equilibrium occurs on the backward-bending portion of the industry supply curve.

Following Proposition 4-4-2 any quota which does not exceed the regulated equilibrium output level consistent with the amount of enforcement does not affect the regulated equilibrium output level. Given that, as in the case of a "low-cost" fishery, the NPIT regulator is indifferent as to the distribution of the rental value of the quota between fines and rents accruing to holders of quota, any quota which does not exceed the regulated equilibrium output level consistent with the NPIT-optimal level of enforcement is optimal for the NPIT regulator.

With a quota which by assumption does not exceed the unregulated open-access equilibrium this condition is satisfied for all regulated equilibria which occur on the backward-bending portion of the penalty-inclusive industry supply curve and for all those which occur on the upward-sloping portion of the penalty-inclusive industry supply curve at output levels that exceed the unregulated open-access equilibrium output level. Given that the quota can be costlessly adjusted by the regulator the condition continues to hold at all regulated equilibria which occur on the upward-sloping portion of the penalty-inclusive industry supply curve at output levels less than the unregulated open-access equilibrium output level.

These results are summarized in the following Proposition.

PROPOSITION 4-5-4: Under the assumptions in the stem of Proposition 4-4-2, the NPIT-optimal regulated equilibrium in the case of a "high-cost" fishery may occur on either the downward-sloping or upward-sloping portions of the penalty-inclusive industry supply curve or at the MSY level of output. Any quota level which does not exceed the regulated equilibrium output level consistent with the NPIT-optimal amount of enforcement is optimal for the NPIT regulator. Given that the output quota is costlessly adjustable by the regulator the characteristics of the NPIT-optimal regulated equilibrium in any particular fishery and its associated output level depend on the levels and rates of change of the marginal cost and marginal benefit of enforcement.

### (4) A Comparison of Regulation in a "Low-Cost" and "High-Cost" Fishery

In the case where the NPIT-optimal regulated equilibrium for a "high-cost" fishery occurs in the upward-sloping portion of the penalty-inclusive industry supply curve, the associated output level will, in general, differ from that which is optimal for the NPIT regulator in a similar "low-cost" fishery.

PROPOSITION 4-5-5: Suppose that two fisheries have identical cost structures but that one fishery is a "low-cost" fishery as for Proposition 4-4-1 while the other is a "high-cost" fishery as for Proposition 4-4-2. Suppose also that at any given quantity the demand price in the "high-cost" fishery exceeds that in the "low-cost" fishery.

(i) In a world of costless enforcement the socially-optimal output level in the "high-cost" fishery exceeds that of the "low-cost" fishery.

(ii) Suppose that enforcement is costly and that the NPIT-optimal regulated equilibrium in the "high-cost" fishery occurs on the upward-sloping portion of the penalty-inclusive industry supply curve. If the marginal productivity of enforcement is similar in both fisheries the NPIT-optimal regulated equilibrium output level in the "high-cost" fishery exceeds that in the "low-cost" fishery.

Proof.

(i) As shown, with reference to Figure 4-3-7, the socially-optimal output level in a world of costless enforcement is that output level which equates marginal social revenue, or demand price, with marginal social cost which in this case is given by the private marginal cost of production. Thus, using (4-5-3) and (4-5-4), at the socially-optimal output level $\hat{Q}$

$$h(\hat{Q},Y) = C_1(\hat{Q},\phi) \tag{4-5-17}$$

Following the assumption in the Proposition concerning relative demand prices

$$h^{HC}(Q(.),Y) > h^{LC}(Q(.),Y) \tag{4-5-18}$$

at any given quantity, where $h^{HC}(.)$ and $h^{LC}(.)$ are the demand prices in the "high-cost" and "low-cost" fisheries respectively. From (4-5-17) and (4-5-18) therefore, at the socially-optimal output level of the "low-cost" fishery $\hat{Q}^{LC}$

$$h^{HC}(\hat{Q}^{LC},Y) > C_1(\hat{Q}^{LC},\phi) \tag{4-5-19}$$

Given the assumptions of (4-5-3) and (4-5-4) concerning the slopes of the demand and marginal cost functions, equality between the demand price and marginal cost in the "high-cost" fishery is reached at some higher output level and thus

$$\hat{Q}^{HC} > \hat{Q}^{LC} \tag{4-5-20}$$

(ii) NPIT-optimal regulation in a world of costly enforcement is determined according to the first-order conditions (4-5-5) and (4-5-6). Given that the quota can be costlessly adjusted to remain binding, the NPIT-optimal output level can be determined from (4-5-5) only.

From (4-5-5) any positive level of enforcement that is optimal for the NPIT regulator equates the marginal benefit of enforcement with its marginal cost. Following Proposition 4-4-2, the attainment of a regulated equilibrium at the MSY in the "high-cost" fishery requires a certain level of enforcement. Further enforcement is necessary to generate a regulated equilibrium on the upward-sloping portion of the penalty-inclusive supply curve. Given this, the amount of enforcement associated with any output level, such that regulated equilibrium occurs on the upward-sloping portion of the penalty-inclusive supply curve in each fishery, is greater for the "high-cost" fishery than for the "low-cost" fishery. Using (4-5-1) then

$$\omega'(L_e^{HC}) \geq \omega'(L_e^{LC}) \tag{4-5-21}$$

where $L_e^{HC}$ and $L_e^{LC}$ are the amounts of enforcement required to generate any such output level in the "high-cost" and "low-cost" fisheries respectively.

Given (4-5-18) and the assumption of identical costs

$$h^{HC}(Q(.),Y)-C_1(Q(.),\phi) > h^{LC}(Q(.),Y)-C_1(Q(.),\phi) \tag{4-5-22}$$

where, by Lemma 4-5-1, both sides of (4-5-22) are negative. Multiplying each side of (4-5-22) by their respective marginal productivity of enforcement ($\partial Q/\partial L_e$) gives the marginal benefit of enforcement at the output level Q in each fishery. Given the assumption of similarity of marginal productivities, that is,

$$\frac{\partial Q}{\partial L_e^{HC}} \approx \frac{\partial Q}{\partial L_e^{LC}} \tag{4-5-23}$$

where ≈ denotes approximate equality, the inequality in (4-5-22) dominates any slightly higher marginal productivity that may exist in the "low-cost" case. At any given quantity therefore, the marginal benefit of enforcement in the "high-cost" fishery is less than that in the "low-cost" fishery. Given this and the comparison of marginal costs in (4-5-21), the concavity of (4-5-2) implies that an output level that is optimal for the NPIT regulator in the "low-cost" fishery necessitates the use of more enforcement than is NPIT-optimal in the case of a "high-cost" fishery. Using the result of Proposition 4-4-1 that $\partial Q/\partial L_e < 0$ when the regulated equilibrium occurs on the upward-sloping portion of the penalty-inclusive supply curve, the result here necessarily follows.

□

## 4-6 THE OBJECTIVES OF THE REGULATOR: CT REGULATION

A captured regulator is motivated by the profits of the regulated industry within the regulatory environment and thus is concerned with the distributional effects of its policies. As in Chapter Three, it is assumed that the industry is required to fund some portion (b) of the resource cost of enforcement. The marginal resource cost to the industry of enforcement is then

$$(4\text{-}6\text{-}1) \qquad b\omega'(L_e) > 0 \quad ; \quad 0 < b < 1$$

In addition to this the industry incurs penalty payments on illegal output. Using (4-4-3) and (4-4-5) and assuming that some portion (a) of fine payments are redistributed to the industry, this cost denoted by $EP_I$ is

$$(4\text{-}6\text{-}2) \quad EP_I = [1-a]\lambda(f,L_e)[Q(L_e,R,Z)-R_I] \; ; \quad 0 < a < 1, \quad R_I \leq R$$

where $R_I$ is the amount of available quota (R) initially allocated to members of the industry and $\lambda(f,L_e)$, as defined in (4-4-5), is the per unit expected penalty which is increasing in both the fine rate and the extent of enforcement. The total cost of enforcement to the industry, denoted by $w(L_e,R_I,R,a,b)$ is

$$(4\text{-}6\text{-}3) \qquad w(L_e,R_I,R,a,b) = b\omega(L_e) + EP_I$$

### (1) Optimal Policy in a "Low-Cost" Fishery

Industry profits in the regulated environment are given by the difference between revenues and the cost of production together with the cost of regulation. The CT regulator is assumed to maximize the value of regulated profits $\pi_{R_e}$. The CT regulator's objective function is then

$$\text{(4-6-4)} \quad \underset{L_e, R_I, R}{\text{Maximize}} \; \pi_{R_e} = h(Q(L_e,R,Z),Y)Q(L_e,R,Z) - C(Q(L_e,R,Z),\phi) - w(L_e,R_I,R,a,b)$$

<u>LEMMA 4-6-1:</u> <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of a constant flat rate per unit expected penalty that is binding on behaviour. A CT regulator ensures that all available quota is allocated to members of the industry.</u>

<u>Proof</u>.

Differentiating (4-6-4) with respect to the amount of available quota allocated to members of the industry, $R_I$, gives

$$\text{(4-6-5)} \quad \frac{\partial \pi_{R_e}}{\partial R_I} = -\frac{\partial w}{\partial R_I} = [1-a]\lambda(f,L_e) > 0$$

substituting from (4-6-3) and (4-6-2). This shows that an increase in the amount of available quota allocated to members of the industry increases their profits in the regulated environment. Hence a captured regulator will always ensure that the maximum possible quota, that is, all available quota, is allocated to members of the industry and thus $R_I = R$.

□

As explained in Section 4-4 the net per unit fine rate $[1-a]\lambda(f,L_e)$ is the trading price of the quota. The initial allocation of quota confers a rent upon the holder equal to this amount. Lemma 4-6-1 therefore shows that a CT regulator, in allocating any given quota, seeks to maximize the rents which accrue to members of the industry from the initial allocation of such quota.

Given that $R_I = R$, CT-optimal regulatory policy can be determined by considering enforcement and aggregate quota only. Differentiating (4-6-4) gives the first-order conditions

$$(4\text{-}6\text{-}6) \quad \frac{\partial \pi_{R_e}}{\partial L_e} = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial Q}{\partial L_e} - \frac{\partial w}{\partial L_e} \leq 0 \;; \quad < \text{ only if } \tilde{L}_e = 0$$

$$(4\text{-}6\text{-}7) \quad \frac{\partial \pi_{R_e}}{\partial R} = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial Q}{\partial R} - \frac{\partial w}{\partial R} \lesseqgtr 0 \;; \quad \begin{array}{l} < \text{ only if } \tilde{R} = 0 \\ > \text{ only if } \tilde{R} = Q(\tilde{L}_e,\tilde{R},Z) \end{array}$$

Assuming that (4-6-4) is concave, (4-6-5) and (4-6-6), when solved simultaneously, generate the CT-optimal level of enforcement $\tilde{L}_e$ and quota size $\tilde{R}$.

<u>PROPOSITION 4-6-1</u>: <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of constant flat rate per unit expected penalty that is binding on behaviour. Assuming that the unregulated open-access equilibrium occurs on the upward sloping portion of the industry supply</u>

curve, that it is optimal for the CT regulator to enforce the quota, and that the level of the quota can be costlessly and instantaneously adjusted by the regulator;

(i) CT-optimal enforcement occurs where the marginal benefit of enforcement to the industry is equal to its marginal cost to the industry. It is not optimal for the CT regulator to restrict output to the pure monopoly profit maximizing level.

(ii) The CT-optimal quota size is equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement as determined in part (i) of the Proposition.

Proof.

The result of part (ii) of the Proposition will be used in the proof of part (i). Part (ii) is therefore proved first.

(ii) Following part (ii) of Proposition 4-4-1, a change in the level of output quota, such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, leaves the regulated equilibrium output level unchanged. Thus $\partial Q/\partial R = 0$ and the first term on the right-hand side of (4-6-7) is also zero.

From (4-6-7) using (4-6-3), (4-6-2) and Lemma 4-6-1

(4-6-8) $$\frac{\partial \pi_{R_e}}{\partial R} = -\frac{\partial w}{\partial R} = [1-a)\lambda(f,L_e) > 0$$

This shows that an increase in quota such that the quota level does not exceed the regulated equilibrium output level consistent with the amount of enforcement acts to increase industry profits in the regulated environment. A captured regulator will therefore maximize the amount of available quota such that this condition holds. Thus the CT-optimal quota will be equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement as determined in part (i) of the Proposition.

(i) The bracketed term in (4-6-6) shows the difference between marginal revenue and marginal production cost in the regulated environment. Given that the unregulated open-access equilibrium exceeds the pure monopoly profit-maximizing level where marginal revenue equals marginal cost, this term is negative in the output region considered. This difference between marginal revenue and marginal cost represents the per unit amount of potential monopoly profit to be realized by regulation. Multiplying this by $\partial Q/\partial L_e$, which by Proposition 4-4-1 is negative, gives the marginal benefit of enforcement to the industry.

The remaining term in (4-6-6) represents the marginal cost of enforcement to the industry. Following part (ii) of the Proposition and Lemma 4-6-1,

(4-6-9) $$\tilde{R}_I = \tilde{R} \ ; \ \tilde{R} = Q(\tilde{L}_e, \tilde{R}, Z)$$

where $\tilde{R}_I$ is the amount of available quota initially allocated to incumbent members of the industry. Given the assumption that it is institutionally costless to adjust the level of available quota, and that any such adjustment can be instantaneously accomplished, profit-maximizing behaviour by the CT regulator ensures that (4-6-9) holds at all times. Using (4-6-9) in (4-6-2) shows that the expected penalty payments incurred by members of the industry are zero and hence the only cost of enforcement to the regulator is the resource cost as given in (4-6-1). Differentiating (4-6-3) with respect to the level of enforcement then gives

(4-6-10) $$\frac{\partial w}{\partial L_e} = b\omega'(L_e) > 0$$

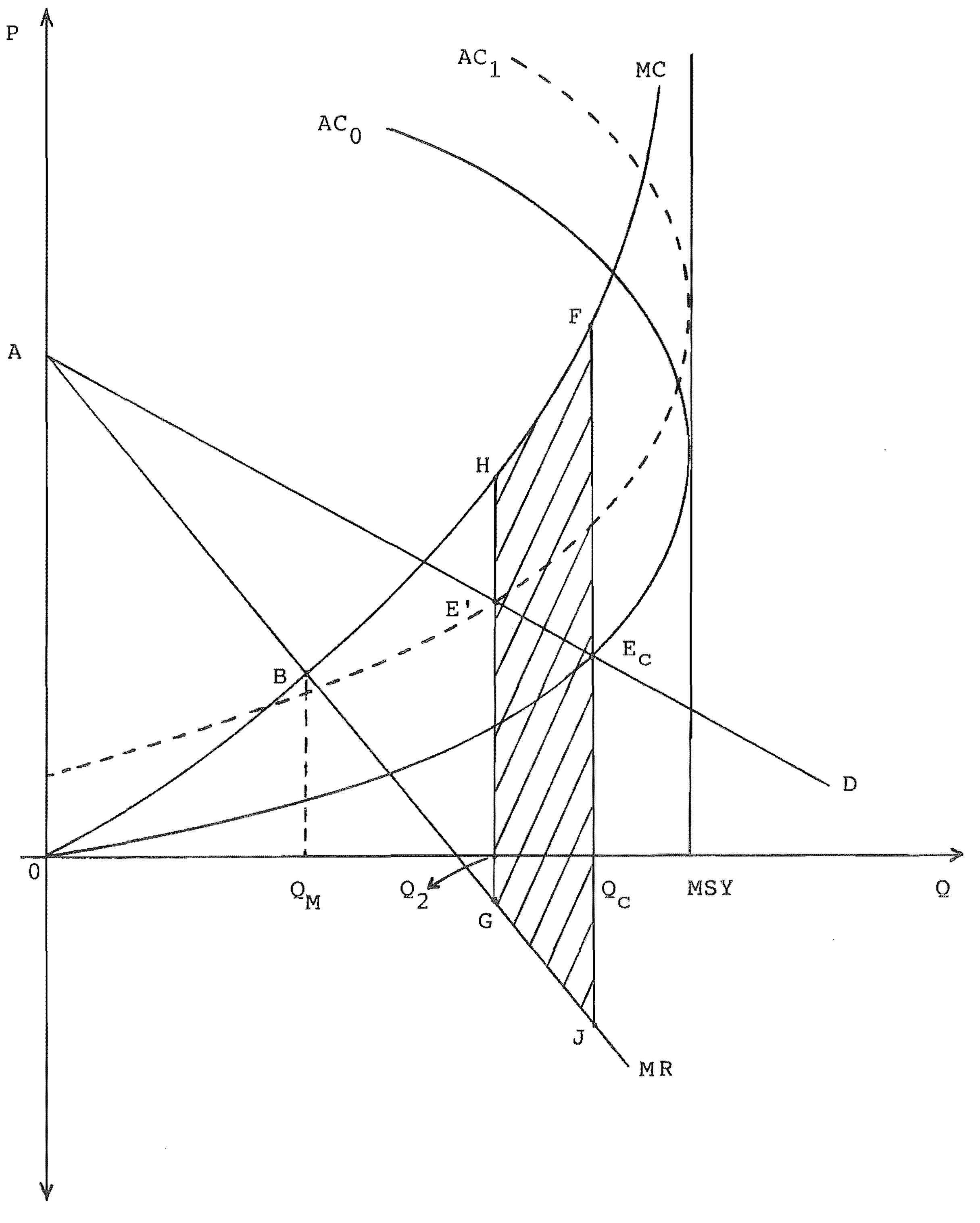
following the assumption in (4-6-1) that the marginal resource cost of enforcement is positive.

Following (4-6-6), CT-optimal enforcement occurs where $\partial\pi_{R_e}/\partial L_e = 0$. Given the interpretations in the preceding paragraphs, this then requires equality between the marginal benefit and marginal cost of enforcement to the industry. At the pure monopoly profit-maximizing output level, marginal revenue equals marginal cost and hence the first term of (4-6-6) is zero. Given the result of (4-6-10) that the marginal cost to the industry of enforcement is positive, it is clearly not optimal for a CT-regulator to restrict the output of the fishery to the pure monopoly profit-maximizing level.

□

The situation of CT regulation of a "low-cost" open-access fishery is illustrated in Figure 4-6-1 below.

Figure 4-6-1: CT regulation of a "low-cost" open-access fishery.



As before, the initial open-access equilibrium is at $E_c$. The imposition of an enforced quota again shifts the supply curve vertically by the amount of the expected penalty. Given that the quota is allocated entirely to members of the industry, the original marginal cost curve can be used to analyse the costs and benefits of regulation to the industry.

The quantity restriction and the resultant price rise unambiguously reduces consumer surplus but that is irrelevant to the objective of the CT regulator. At $E_c$, the profits of 0AB made on the initial units are cancelled out by the loss of BFJ resulting in zero profit. The reduction in output to $Q_2$ results in a net gain to the producers of the shaded area HFJG.

(2) Policy Responses to Parameter Changes in a "Low-Cost" Fishery

Following Proposition 4-6-1, the CT-optimal regulated equilibrium output level, $\tilde{Q} = Q(\tilde{L}_e, \tilde{R}, Z)$, occurs at some intermediate level

$$(4\text{-}6\text{-}11) \qquad Q_M < \tilde{Q} < Q_c \; ; \quad \tilde{L}_e > 0, \; \tilde{R} = \tilde{Q}$$

The effects of parameter changes on the CT-optimal regulatory policy associated with this regulated equilibrium can be examined by totally differentiating the first order conditions (4-6-6) and (4-6-7). Totally differentiating (4-6-6) gives

$$(4\text{-}6\text{-}12) \quad A d\tilde{L}_e + B d\tilde{R} + E dY + J d\phi + S da + T db = 0$$

where

$$A = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e^2} - \frac{\partial^2 w}{\partial L_e^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

$$B = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e \partial R} - \frac{\partial^2 w}{\partial L_e \partial R}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial R}\right]\frac{\partial Q}{\partial L_e}$$

$$E = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e \partial Y} - \frac{\partial^2 w}{\partial L_e \partial Y}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial Y}\right.$$

$$\left. + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)\right]\frac{\partial Q}{\partial L_e}$$

$$J = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e \partial \phi} - \frac{\partial^2 w}{\partial L_e \partial \phi}$$

$$+ \left[[h_{11}(Q(.),Y)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial \phi}\right.$$

$$\left. - C_{12}(Q(.),\phi)\right]\frac{\partial Q}{\partial L_e}$$

$$S = -\frac{\partial^2 w}{\partial L_e \partial a}$$

$$T = -\frac{\partial^2 w}{\partial L_e \partial b}$$

and totally differentiating (4-6-7) gives,

$$(4\text{-}6\text{-}13) \quad Md\tilde{L}_e + Nd\tilde{R} + FdY + Kd\phi + Uda + Vdb = 0$$

where

$$M = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial R \partial L_e} - \frac{\partial^2 w}{\partial R \partial L_e}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial R}$$

$$N = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial R^2} - \frac{\partial^2 w}{\partial R^2}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial R}\right]\frac{\partial Q}{\partial R}$$

$$F = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial R \partial Y} - \frac{\partial^2 w}{\partial R \partial Y}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial R}\right.$$

$$\left. + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)\right]\frac{\partial Q}{\partial R}$$

$$K = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial R \partial \phi} - \frac{\partial^2 w}{\partial R \partial \phi}$$

$$+ \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial \phi}\right.$$

$$\left. - C_{12}(Q(.),\phi)\right]\frac{\partial Q}{\partial R}$$

$$U = -\frac{\partial^2 w}{\partial R \partial a}$$

$$V = -\frac{\partial^2 w}{\partial R \partial b}$$

The preceding analysis has shown that, at any given vector of parameter values, it is always optimal for the CT regulator to set the quota equal to the regulated equilibrium output level consistent with the optimal level of enforcement. In this situation no quota violations occur and no fines are incurred by the industry. It is not optimal for the regulator to enforce a quota which exceeds the regulated equilibrium output level consistent with the level of enforcement because in that case the marginal productivity of enforcement is zero. Neither is it optimal to have a quota which is smaller than the regulated equilibrium output level consistent with the optimal level of enforcement for in that case the industry incurs a net loss of profits through the payment of fines.

Given that the CT-optimal quota is set so that no violations occur, and assuming that the quota can be costlessly and instantaneously adjusted by the regulator, this condition continues to hold in the case of parameter-induced changes in CT-optimal regulatory policy. The optimal level of enforcement then is the only freely determined policy variable while the quota level is automatically adjusted by the regulator to remain consistent with the condition required for optimality given in part (ii) of Proposition 4-6-1.

The first parameter change to be considered concerns the level of aggregate consumer income.

(a) <u>The level of aggregate consumer income</u>

<u>PROPOSITION 4-6-2:</u> <u>Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota enforced by means of a constant flat rate per unit expected penalty that is binding on behaviour. Suppose also that the unregulated open-access equilibrium occurs on the positively sloped portion of the industry supply curve, that the quota, being</u>

costlessly and instantaneously adjustable, is always set at the regulated equilibrium level consistent with the amount of enforcement so that no violations occur, that the CT regulator's objective function $\pi_{R_e}$ is concave in enforcement, and that the CT-optimal regulatory policy at some fixed vector of parameters involves strictly positive levels of enforcement and quota in accordance with Proposition 4-6-1. Assuming that a change in aggregate income leaves the marginal productivity of the regulator's policy instruments unchanged, an increase in aggregate consumer income, which increases demand for the output of the fishery, reduces (increases) the CT-optimal level of enforcement if the combined effect of the income-induced increase in demand price and change in the price responsiveness of demand, evaluated at the initial regulated equilibrium output level, exceeds (is less than) the income-induced change in the per unit loss of industry profit on the marginal unit of output. Ceteris paribus, the increase in income is more likely to lead to an increase in enforcement if it reduces

<u>the price responsiveness of demand than if it increases it.</u>

<u>Proof</u>.

Recalling, from part (ii) of Proposition 4-4-1, that a change in the level of quota, such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, leaves the regulated equilibrium output level unchanged so that $\partial Q/\partial R = 0$, it follows in (4-6-12) and (4-6-13) that $\partial^2 Q/\partial R \partial L_e = \partial^2 Q/\partial L_e \partial R = \partial^2 Q/\partial R \partial Y = \partial^2 Q/\partial R^2 = 0$.

Given that the quota is continuously set at the regulated equilibrium output level so that no violations occur it follows from (4-6-2) and (4-6-3) that the only cost of enforcement to the regulator is the resource cost and therefore $\partial w/\partial R = \partial^2 w/\partial R \partial L_e = \partial^2 w/\partial R \partial Y = \partial^2 w/\partial R^2 = 0$.

When the above conditions are substituted in (4-6-13) the terms in M, N and F are all zero. Therefore, using (4-6-12) and (4-6-13) with $d\phi = da = db = 0$ shows that

$$(4\text{-}6\text{-}14) \qquad \frac{d\tilde{L}_e}{dY} = -\frac{E}{A} \gtrless 0 \quad \text{if and only if } E \gtrless 0$$

given that $A < 0$ from the concavity in enforcement of (4-6-4).

Following the assumption that a change in aggregate consumer income does not affect the marginal productivity of the regulator's policy instruments, the component term $\partial^2 Q/\partial L_e \partial Y$ in (4-6-12) is zero.

From (4-6-12) then

$$(4\text{-}6\text{-}15) \quad E = \Big[ [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial Y} + h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y)\Big]\frac{\partial Q}{\partial L_e} - \frac{\partial^2 w}{\partial L_e \partial Y}$$

Rearranging (4-6-15), $E \gtrless 0$ if and only if

$$(4\text{-}6\text{-}16) \quad h_{12}(Q(.),Y)Q(.)+h_2(Q(.),Y) \lesseqgtr \frac{\partial^2 w}{\partial L_e \partial Y} / \frac{\partial Q}{\partial L_e} - [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial Y}$$

Following Appendix 4-1, an increase in income, which increases demand for the output of the fishery, acts to increase the regulated equilibrium output of the fishery and thus $\partial Q/\partial Y > 0$. From (A4-2-3) it is assumed that $[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C_{11}(Q(.),\phi)] < 0$ while, following (4-6-10), the marginal cost of enforcement to the industry comprises only the marginal resource cost of such enforcement which is unaffected by changes in the level of aggregate consumer income and hence $\partial^2 w/\partial L_e \partial Y$ equals zero. Given these results, and those of previous discussion, the right-hand side of (4-6-16) is some positive number.

The square bracketed term on the right-hand side of (4-6-16) shows the unit rate of change in the difference between marginal revenue and marginal cost. Given that, following the proof of Proposition 4-6-1, marginal revenue is less than marginal cost over the relevant output range, multiplying this square bracketed term by $\partial Q/\partial Y$ gives the income-induced change in the unit loss of industry profit on the marginal unit of output.

On the left-hand side of (4-6-16), the term $h_2(Q(.),Y)$ denotes the income-induced change in demand price at the initial regulated equilibrium quantity which, following the assumption of the Proposition, is positive. As previously derived, the effect of a change in income on the slope of the demand curve is given by $h_{12}(Q(.),Y)$. If this term is negative (positive), an increase in income increases (reduces) the absolute value of the slope of the demand curve or alternatively reduces (increases) the price responsiveness of demand.

Given the above results and definitions, using (4-6-16) in (4-6-14) reveals the validity of the Proposition. From (4-6-14), the increase in income leads to an increase in CT-optimal enforcement and a concomitant decrease in the optimal quota level if and only if $F > 0$. Following (4-6-16) this inequality is more likely to hold if $h_{12}(Q(.),Y) < 0$ than if $h_{12}(Q(.),Y) > 0$. Ceteris paribus, therefore, the increase in income is more likely to lead to an increase in the level of CT-optimal enforcement if it reduces the price responsiveness of demand than if it increases it.

□

The result of this Proposition is similar to that of part (i) of Proposition 3-5-1 as is its interpretation. Recalling that $\partial^2 w/\partial L_e \partial Y = 0$, multiplying the terms of (4-6-16) by $\partial Q/\partial L_e$ reveals that the response in CT-optimal enforcement to the change in income depends on the income-induced changes in the marginal benefit and marginal cost of enforcement to the regulator. Again, the result concerning the income-induced change in the price responsiveness of demand is motivated by the fact that the CT regulator is essentially acting as a monopolist in restricting output and that the less price responsive is demand, the greater is the scope for profitably exercising monopoly power.

Following part (ii) of Proposition 4-6-1 the CT-optimal quota is set at the regulated equilibrium output level generated by the new CT-optimal level of enforcement at the new level of aggregate income. Of itself, as shown in Appendix 4-1, an increase in aggregate income increases the competitive equilibrium output level of a low-cost fishery and hence also the regulated equilibrium output level at any level of enforcement. If as a result of an increase in income it is optimal for the regulator to reduce enforcement activity the regulated equilibrium output level unambiguously increases and hence so too does the CT-optimal quota to match. If, however, it is optimal for the regulator to increase enforcement activity in response to an increase in aggregate income the effect on the regulated equilibrium output level and hence the CT-optimal quota level is ambiguous and depends on the precise functional forms of the demand and cost curves for the fishery.

The second parameter change considered concerns the structure of marginal production cost in the fishery. Here, as in Proposition 4-5-3 for the case of NPIT regulation, this is proxied by the parameter $\phi$ which may represent the state of available technology or the mortality rate of the fishery.

(b) <u>The structure of marginal production cost</u>

<u>PROPOSITION 4-6-3</u>: <u>Under the assumptions in the stem of Proposition 4-6-2</u>;

(i) A change in a parameter which leads to an increased marginal cost of production at any output level, but which does not affect the marginal productivity of enforcement, reduces (increases) the CT-optimal level of enforcement if and only if the parameter-induced change in the marginal cost of production at the initial regulated equilibrium is less than (exceeds) the parameter-induced change in the per unit loss of industry profit on the marginal unit of output.

(ii) Assuming that a change in a parameter which increases the marginal cost of production at any output level also improves (reduces) the marginal productivity of enforcement, if the parameter-induced change in the marginal cost of production at the initial regulated equilibrium exceeds (is less than) the parameter-induced change in the per unit loss of industry profit on the marginal unit of output, the change in the parameter increases (reduces) the

CT-optimal level of enforcement. If, however, the parameter-induced change in the marginal cost of production at the initial regulated equilibrium is less than (exceeds) the parameter-induced change in the per unit loss of industry profit on the marginal unit of output, the effect of the parameter change on CT-optimal regulatory policy is ambiguous.

Proof.

(i) Using (4-6-12) with $dY = da = db = 0$ gives

(4-6-17) $$\frac{d\tilde{L}_e}{d\phi} = -\frac{J}{A} \gtrless 0 \quad \text{if and only if} \quad J \gtrless 0$$

given that $A < 0$ by the concavity in enforcement of (4-6-4) and that, as derived in the proof of Proposition 4-6-2, $B = 0$.

Following the assumption that the parameter change does not affect the marginal productivity of enforcement, $\partial^2 Q/\partial L_e \partial\phi = 0$ and, from (4-6-10), $\partial^2 w/\partial L_e \partial\phi$ is also equal to zero. From (4-6-13) then

(4-6-18) $$J = \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial\phi} - C_{12}(Q(.),\phi)\right]\frac{\partial Q}{\partial L_e}$$

Rearranging (4-6-18), $J \gtrless 0$ if and only if

(4-6-19) $$C_{12}(Q(.),Y) \gtrless [h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y) - C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial\phi}$$

From the discussion in Appendix 4-1, $\partial Q/\partial\phi < 0$ while, following (A4-2-3), it is assumed that

$[h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C_{11}(Q(.),\phi)] < 0.$

The right-hand side of (4-6-19) is therefore some positive number. The interpretation of the square bracketed term is as in the proof of Proposition 4-6-2. Multiplied by $\partial Q/\partial\phi$, it gives the parameter-induced change in the per unit loss of industry profit on the marginal unit of output. The term $C_{12}(Q(.),\phi)$ on the left-hand side of (4-6-19) denotes, following (4-5-4), the parameter-induced change in marginal production cost at the initial regulated equilibrium. Given these results and interpretations, applying (4-6-19) in (4-6-17) reveals the validity of the Proposition.

(ii) If the parameter change affects the marginal productivity of enforcement, the term in $\partial^2Q/\partial L_e\partial\phi$ is non zero. Manipulating J from (4-6-12) reveals that this results in the subtraction of the term

$$[h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e\partial\phi}/\frac{\partial Q}{\partial L_e} \qquad (4\text{-}6\text{-}20)$$

from the left-hand side of (4-6-19).

Given the result that marginal revenue is less than marginal cost in the relevant output range and that $\partial Q/\partial L_e < 0$, the sign of the term in (4-6-20) varies directly with that of $\partial^2Q/\partial L_e\partial\phi$. If the parameter change serves to increase the marginal productivity of enforcement $\partial^2Q/\partial L_e\partial\phi < 0$ and hence (4-6-20) is negative. From the above argument, the left-hand side of (4-6-19) becomes more positive. If, in the absence of this effect, the "greater than" inequality prevailed in (4-6-19), so that the effect of the parameter change was to increase CT-optimal enforcement, the inclusion of a parameter-induced improvement in the marginal productivity of enforcement reinforces the result. Alternatively, if the opposite inequality prevailed in (4-6-19) so that the effect of the paramater change, in the absence of any parameter-induced change in the marginal productivity of enforcement, was to reduce CT-optimal enforcement, the inclusion of such a modification would create ambiguity in the result. The validity of the Proposition in the opposite case where the parameter change reduces the marginal productivity of enforcement can be established using the same argument with the inequalities reversed.

□

The results here concerning the parameter $\phi$ are the converse of those of Proposition 4-6-2 relating to a change in the level of aggregate consumer income. An increase in $\phi$ which raises marginal cost reduces the pure monopoly profit-maximizing output level of the fishery from its initial level as shown by $Q_M$ in Figure 4-6-1. This by itself would lead to an increase in the marginal benefit of enforcement at any output level exceeding $Q_M$ and thus stimulate an increase in CT-optimal enforcement activity. However, the increased cost structure also leads to a reduction in the regulated equilibrium output level of the fishery at any level of enforcement. This reduces the loss of industry profit on the marginal unit of output thus reducing the marginal benefit of enforcement. The CT-optimal policy response is determined by a comparison of the magnitudes of these two effects. As evidenced by part (ii) of the Proposition, parameter-induced changes in the marginal productivity of enforcement may reinforce or reverse the results in part (i) of the Proposition depending on their sign and relative size.

As stated above, an increase in the cost structure of a "low-cost" fishery reduces the competitive equilibrium output level of the fishery and hence also the regulated equilibrium output level at any level of enforcement. If, as a result of the increase in the cost structure, it is optimal for the regulator to increase enforcement activity, the regulated equilibrium output level is therefore unambiguously reduced and hence, following (4-6-9), so too is the CT-optimal quota. If, however, it is optimal for the regulator to reduce enforcement activity in response to an increase in the cost structure parameter $\phi$, the effect on the regulated equilibrium output level and hence the CT-optimal quota level is ambiguous and, as in the case of the aggregate income parameter, depends on the exact functional forms of the supply and demand curves of the fishery.

The final parameter changes examined here are the distributional parameters concerning the redistribution of fine revenues and the proportional funding requirements of the resource cost of enforcement.

(c) Changes in distributional parameters

PROPOSITION 4-6-4: Under the assumptions contained in the stem of Proposition 4-6-2;

(i) A change in the proportion of fine payments that are redistributed to members of the industry has no effect on the CT-optimal levels of enforcement and quota.

(ii) An increase in the proportion of the resource cost of enforcement that the industry is required to fund leads to a reduction in the CT-optimal enforcement level and a concomitant increase in the CT-optimal quota.

Proof.

(i) Using (4-6-2) with $dY = d\phi = db = 0$ gives

$$\frac{d\tilde{L}_e}{da} = -\frac{S}{A} \gtreqless 0 \quad \text{if and only if} \quad S \gtreqless 0 \tag{4-6-21}$$

given that $A < 0$ by the concavity in enforcement of (4-6-4) and that, as previously derived, $B = 0$.

Following the assumption that the quota is continuously set so that no violations occur, in (4-6-3), the only cost of enforcement to the regulator is the resource cost. This is unaffected by changes in the parameter 'a' and hence, from (4-6-10), $\partial^2 w/\partial L_e \partial a = 0$. Using this result in (4-6-21), given the definition of S from (4-6-12), reveals that $d\tilde{L}_e/da = 0$. Given that the CT-optimal level of enforcement is unchanged, the regulated equilibrium output level remains unaltered and so too, following (4-6-9), does the optimal level of quota. CT-optimal regulatory policy is therefore unaffected by changes in the parameter 'a'.

(ii) Using (4-6-12) with $dY = d\phi = da = 0$ gives

$$\frac{d\tilde{L}_e}{db} = -\frac{T}{A} \gtrless 0 \quad \text{if and only if} \quad T \gtrless 0 \tag{4-6-22}$$

given the results on terms A and B as in part (i) of the Proposition.

From (4-6-10), which follows from (4-6-3) given the assumption on quota size,

$$\frac{\partial^2 w}{\partial L_e \partial b} = \omega'(L_e) > 0 \tag{4-6-23}$$

Using this result in (4-6-22), given the definition of T from (4-6-12), reveals that $d\tilde{L}_e/db < 0$ and hence that an increase in the proportion of the resource cost of enforcement that the industry is required to pay reduces the CT-optimal level of enforcement activity. Following Proposition 4-4-1 this parameter-induced reduction in enforcement results in an increase in the regulated equilibrium output level and also, given (4-6-9), in a matching rise in the CT-optimal level of quota. Hence an increase in the proportion of the resource cost of enforcement that the industry is required to pay causes a relaxation in CT-optimal regulatory policy for the fishery.

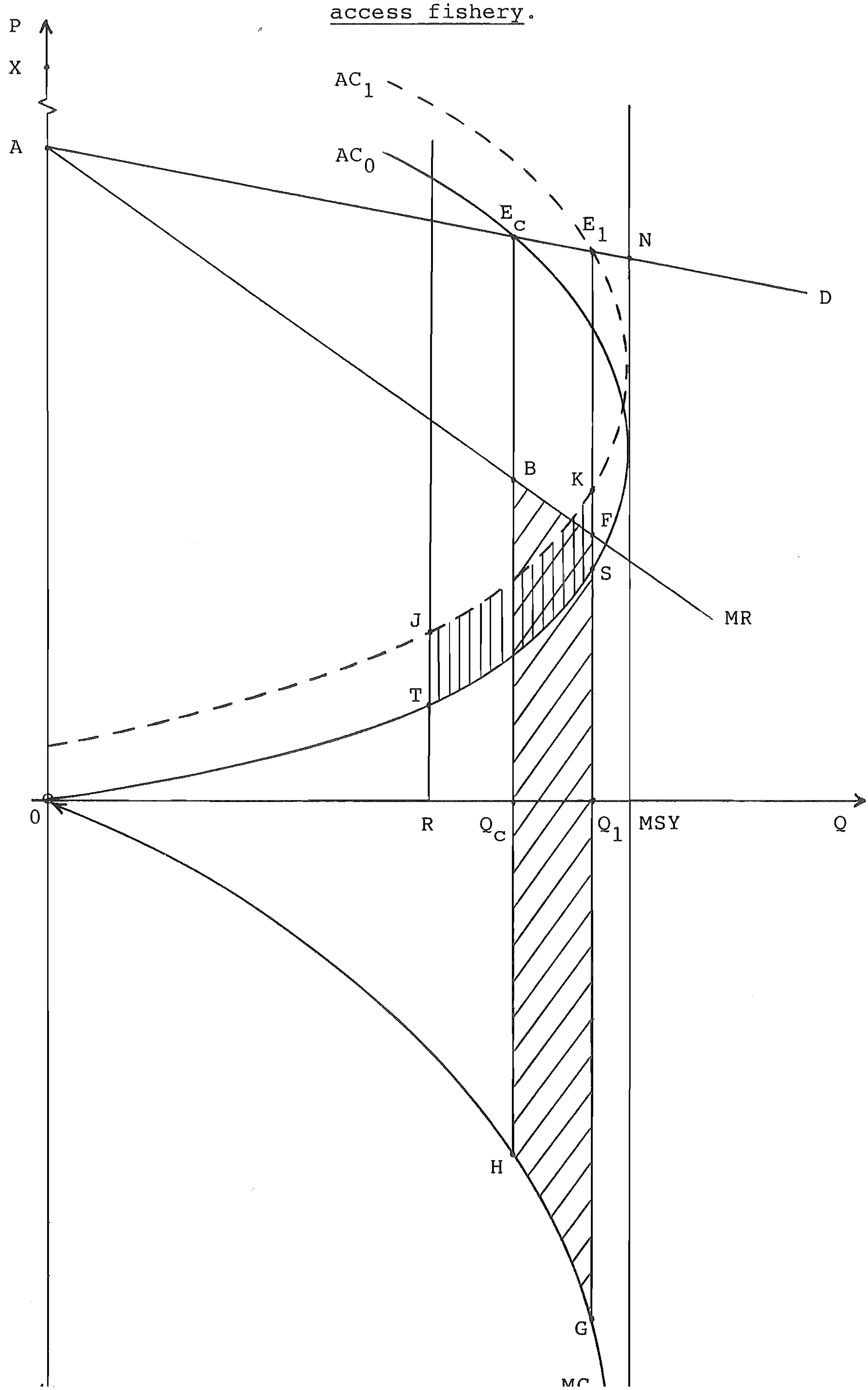
□

Proposition 4-6-4 is intuitively obvious. The result that a ceteris paribus increase in the marginal cost of enforcement to the industry reduces CT-optimal enforcement follows immediately while result (i) stems from the fact that a captured regulator ensures that there are no fine revenues to be redistributed.

(3) Optimal Policy in a "High-Cost" Fishery

These results refer to the CT regulation of a "low-cost" fishery. In the case of a "high-cost" fishery, illustrated in Figure 4-6-2, the objective of the regulator remains the same; to maximize industry profit as expressed in (4-6-4). However, at all regulated equilibria which

Figure 4-6-2: CT regulation of a "high-cost" open-access fishery.



occur on the backward-bending portion of the penalty-inclusive industry average cost curve, marginal revenue exceeds marginal cost, which is negative, and, following Proposition 4-4-2, $\partial Q/\partial L_e > 0$.

At the initial competitive equilibrium output $Q_c$ in Figure 4-6-2 producer surplus is zero. Marginal revenue, however, exceeds marginal cost and therefore industry output can be profitably expanded. Given the dynamics of the fishery as outlined in Section 4-3, this expansion can be accomplished through the imposition and enforcement of a binding output quota. Figure 4-6-2 illustrates such a quota set at some level $R < Q_c$ and enforced by a marginal expected penalty of magnitude JT = KS. This generates a regulated equilibrium at $E_1$ with output $Q_1$ and results in an increase in industry profit of the diagonally shaded area BFGH. Given, however, that the quota must initially be set at some output level less than that at the unregulated open-access equilibrium in order to affect behaviour and is constrained by assumption not to exceed the unregulated equilibrium output level, the industry will incur expected penalty payments on illegal output. Allowing for a proportion $0 < a < 1$ of these payments to be refunded to members of the industry, net industry profit will be smaller than that described above by an amount $(1 - a)$ times the vertically shaded area JKST.

The objective function of the CT regulator (4-6-4) in the case of a "high-cost" fishery is no longer unambiguously concave. The marginal resource cost of enforcement to the industry is unaffected by the distinction between "high-cost" and "low-cost" fisheries. The first term of the first-order condition (4-6-5) again shows the marginal benefit of enforcement. As shown in Figure 4-6-2, marginal cost is negative in the region for which the supply curve is negatively sloped and, as stated above, at regulated equilibria which occur on the backward-bending portion of the long run penalty-inclusive supply curve,

$$(4\text{-}6\text{-}24)\quad h_1(Q(.),Y)Q(.) + h(Q(.),Y) - C_1(Q(.),\phi) > 0$$

This gives the per unit amount of potential industry profit. Multiplied by $\partial Q/\partial L_e$, which from Proposition 4-4-2 is positive, it represents the marginal benefit of enforcement.

The fine payments incurred by the industry in a "high-cost" fishery represent an additional complication. Using (4-6-2) and (4-6-3)

$$(4\text{-}6\text{-}25)\quad \frac{\partial w}{\partial L_e} = b\omega'(L_e)+(1-a)[\lambda(f,L_e)\frac{\partial Q}{\partial L_e}+[Q(.)-R]\frac{\partial \lambda}{\partial L_e}] > 0$$

given that, as previously defined, all component terms are positive. Differentiating (4-6-5) with respect to enforcement, using (4-6-25) gives

$$(4\text{-}6\text{-}26)\quad \frac{\partial^2 \pi_{R_e}}{\partial L_e^2} = \Big[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)$$

$$-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial L_e}\Big]\frac{\partial Q}{\partial L_e}+[h_1(Q(.),Y)Q(.)+h(Q(.),Y)$$

$$-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e^2}+b\,\omega''(L_e)+(1-a)[\lambda(f,L_e)\frac{\partial^2 Q}{\partial L_e^2}$$

$$+2\frac{\partial \lambda}{\partial L_e}\frac{\partial Q}{\partial L_e} + [Q(.)-R]\frac{\partial^2 \lambda}{\partial L_e^2}]$$

Depending on the relative magnitudes of the first two component terms of (4-6-26) the marginal profitability of enforcement may be increasing or decreasing as the level of enforcement rises. As previously stated, at the MSY marginal production cost is undefined while $\partial Q/\partial L_e = 0$. Together these two properties imply that the marginal profitability of enforcement approaches some finite limit value at that level of enforcement which generates a regulated equilibrium at N in Figure 4-6-2 where output is at the MSY level. If this limit value exceeds the marginal cost of the associated level of enforcement, further enforcement will be undertaken by a CT regulator. In this case, CT-optimal regulatory policy proceeds as in Proposition 4-6-1 for the case of a "low-cost" fishery, and the regulated equilibrium will occur on the upward-sloping portion of the penalty-inclusive supply curve. If, however, the marginal benefit of enforcement is decreasing, or increasing at a slower rate than the marginal cost to the industry of enforcement, it is possible that the CT-optimal regulated equilibrium occurs on the backward-bending portion of the penalty-inclusive industry supply curve.

The interpretation of the first-order condition with respect to quota size (4-6-7) is as for the "low-cost" fishery. Following part (ii) of Proposition 4-4-1, a change in the level of output quota, such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement, leaves the marginal expected penalty, and hence the regulated equilibrium output level, unchanged. In the case when the regulated equilibrium occurs at some output level which exceeds the unregulated open-access equilibrium level, while, by assumption, the quota is constrained not to exceed the unregulated open-access equilibrium output level, industry profit in the regulated environment, as shown in (4-6-8), increases with the size of the output quota. The CT-optimal quota is therefore set at the unregulated open-access equilibrium output level.[10] If, however, it is optimal for the CT regulator to enforce the quota to such an extent that the regulated equilibrium occurs on the upward-sloping portion of the penalty-inclusive industry supply curve at an output level less than the unregulated open-access equilibrium level, the analysis proceeds as for part (ii) of Proposition 4-6-1 and the CT-optimal quota is equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement.

These results are summarized in the following Proposition.

PROPOSITION 4-6-5: Under the assumptions in the stem of Proposition 4-4-2, the CT-optimal regulated equilibrium in the case of a "high-cost" fishery may occur on either the downward-sloping or upward-sloping portions of the penalty-inclusive industry supply curve or at the MSY output level. The characteristics of the CT-optimal regulated equilibrium in any particular fishery and its associated output level depend on the level and rates of change of the marginal cost and marginal benefit to the industry of enforcement. If the regulated equilibrium occurs at some output level which exceeds the unregulated open-access equilibrium level the CT-optimal quota is set at the unregulated open-access equilibrium output level. If, however, the regulated equilibrium occurs on the upward sloping portion of the penalty-inclusive industry supply curve at some output level less than the unregulated open-access equilibrium level the CT-optimal quota is equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement.

(4) A Comparison of Regulation in a "Low-Cost" and "High-Cost" Fishery

In the case where the CT-optimal regulated equilibrium for a "high-cost" fishery occurs on the upward-sloping portion of the penalty-inclusive supply curve, the associated output level will, in general, differ from that which is optimal for a captured regulator in a similar "low-cost" fishery.

PROPOSITION 4-6-6: Suppose that two fisheries have identical cost structures but that one fishery is a "low-cost" fishery as for Proposition 4-4-1 while the other is a "high-cost" fishery as for Proposition 4-4-2. Suppose also that at any given quantity the demand price, and also the marginal revenue in the "high-cost" fishery exceeds that in the "low-cost" fishery.

(i) In a world of costless enforcement the pure monopoly profit-maximizing output level in the "high-cost" fishery exceeds that of the "low-cost" fishery.

(ii) Suppose that enforcement is costly and that the CT-optimal regulated equilibrium in the "high-cost" fishery occurs on the upward-sloping portion of the penalty-inclusive industry supply curve. If the marginal productivity of enforcement is similar in both fisheries, the CT-optimal regulated equilibrium output level in the "high-cost" fishery exceeds that in the "low-cost" fishery.

Proof.

(i) As shown, with reference to Figure 4-3-7, the pure monopoly profit-maximizing output level in a world of costless enforcement is that output level which equates marginal revenue with the marginal cost of production. Thus, at the pure monopoly profit-maximizing output level $Q_M$

(4-6-27) $$h_1(Q(.),Y)Q(.) + h(Q(.),Y) = C_1(Q(.),\phi)$$

Following the assumption in the Proposition concerning relative marginal revenues at any given quantity

(4-6-28) $$[h_1(Q(.),Y)Q(.) + h(Q(.),Y)]^{HC} > [h_1(Q(.),Y)Q(.) + h(Q(.),Y)]^{LC}$$

where the left-hand side of (4-6-28) and the right-hand side of (4-6-28) are the marginal revenues in the "high-cost" and "low-cost" fisheries respectively. From (4-6-27) and (4-6-28) therefore, at the pure monopoly profit-maximizing output level of the "low-cost" fishery $Q_M^{LC}$

(4-6-29) $$[h_1(Q_M^{LC},Y)Q_M^{LC} + h(Q_M^{LC},Y)]^{HC} > C_1(Q_M^{LC},\phi)$$

Given the assumptions of (4-5-3) and (4-5-4) concerning the slopes of the demand and marginal cost functions, equality between marginal revenue and marginal cost in the "high-cost" fishery is reached at some higher output level and thus

(4-6-30) $$Q_M^{HC} > Q_M^{LC}$$

(ii) CT-optimal regulation in a world of costly enforcement is determined according to the first-order conditions (4-6-5) and (4-6-6). Following Proposition 4-4-1, at regulated equilibria occurring on the upward-sloping portion of the penalty-inclusive supply curve, an increase in enforcement has the effect of reducing the regulated equilibrium output of the fishery. If the regulated equilibrium of the "high-cost" fishery occurs at some output level which is less than the unregulated open-access equilibrium level the CT-optimal quota, following Proposition 4-6-5, is equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement. In this case, as in the case of a "low-cost" fishery, following Proposition 4-6-1, the CT-optimal amount of enforcement, and hence the CT-optimal regulated equilibrium output level, can be determined from (4-6-5) only.

From (4-6-5), at any positive level of enforcement that is optimal for the CT regulator, the marginal benefit and marginal cost of enforcement are equated. Following Proposition 4-4-2, generating a regulated equilibrium at the MSY in the "high-cost" fishery requires a certain level of enforcement. Additional enforcement is necessary to generate a regulated equilibrium on the upward-sloping portion of the penalty-inclusive supply curve. Given this, the amount of enforcement required at any output level, such that a regulated equilibrium occurs on the upward-sloping portion of the penalty-inclusive supply curve in each fishery, is greater for the "high-cost" fishery than for the "low-cost" fishery. Using (4-6-3) and (4-5-1) then, recalling that in this instance, the only cost of enforcement to the industry is the proportion of the resource cost that it is required to fund,

$$w'(L_e^{HC}) \geq w'(L_e^{LC}) \tag{4-6-31}$$

where $L_e^{HC}$ and $L_e^{LC}$ are the amounts of enforcement required to generate any such regulated equilibrium output level in the "high-cost" and "low-cost" fisheries respectively.

Given (4-6-28) and the assumption of identical costs

$$\begin{aligned} [h_1(Q(.),Y)Q(.) + h(Q(.),Y)]^{HC} - C_1(Q(.),\phi) \\ > [h_1(Q(.),Y)Q(.) + h(Q(.),Y)^{LC} - C_1(Q(.),\phi) \end{aligned} \tag{4-6-32}$$

where, following Proposition 4-6-1, both sides of (4-6-32) are negative. Multiplying each side of (4-6-32) by their respective marginal productivity of enforcement gives the marginal benefit of enforcement at the output level Q(.) in each fishery. Given the assumption of similarity of marginal productivities,

(4-6-33)
$$\frac{\partial Q}{\partial L_e^{LC}} \approx \frac{\partial Q}{\partial L_e^{HC}}$$

the inequality in (4-6-32) dominates any slightly higher marginal productivity that may exist in the "low-cost" case. At any given quantity therefore, the marginal benefit of enforcement in the "high-cost" fishery is less than that in the "low-cost" fishery. Given this, and the comparison of marginal costs in (4-6-31), the local concavity in enforcement of (4-6-4) implies that the generation of an output level that is optimal for the CT regulator in the "low-cost" fishery case requires a greater level of enforcement than is CT-optimal in the case of a "high-cost" fishery. Using the result from Proposition 4-4-1 that $\partial Q/\partial L_e < 0$ at regulated equilibria which occur on the upward-sloping portion of the penalty-inclusive supply curve, the result necessarily follows. In the case where the regulated equilibrium in the "high-cost" fishery occurs on the upward-sloping portion of the penalty-inclusive industry supply curve but at an output level which exceeds the unregulated open-access equilibrium level the result holds by definition.

□

## 4-7 A COMPARISON OF CT AND NPIT REGULATION

Sections 4-5 and 4-6 presented an analysis of optimal regulatory behaviour under two conflicting hypotheses concerning regulatory objectives. It is evident that the behavioural implications resulting from these hypotheses differ in several respects, notably in the optimal amount of enforcement, the level and distribution of the quota, and the response of policy to changes in the value of particular parameters. The first difference to be considered arises in the optimal level of enforcement.

PROPOSITION 4-7-1: Suppose that an open-access fishery which faces a negatively sloped market demand curve is subject to a binding output quota which can be instantaneously and costlessly adjusted by the regulator and which is enforced by means of a constant flat rate per unit expected penalty that is binding on behaviour. Assuming that the unregulated open-access equilibrium occurs on the upward-sloping portion of the industry supply curve, that enforcing the regulation necessitates the use of scarce resources and incurs a non-decreasing marginal resource cost, and that enforcement is optimal for both regulators, the CT-optimal level of enforcement ($\tilde{L}_e$) exceeds the NPIT-

<u>optimal level ($L_e^*$) and hence the CT-optimal regulated equilibrium output level is lower than that which is optimal for the NPIT regulator.</u>

<u>Proof</u>.

Following Proposition 4-4-1, the regulated equilibrium is unaffected by changes in quota such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement. From Proposition 4-5-1 any quota which does not exceed the regulated equilibrium output level consistent with the NPIT-optimal amount of enforcement is optimal for the NPIT regulator while, from Proposition 4-6-1, it is optimal for the CT regulator to continuously maintain the quota at the regulated equilibrium output level consistent with the CT-optimal amount of enforcement. For the NPIT regulator then the size of the quota within a certain range is irrelevant while for the CT regulator the quota must be continuously maintained at the level such that no penalty is incurred. In each case the optimal quota is passively determined by the setting of the optimal enforcement level in accordance with the appropriate first-order condition with respect to enforcement and a comparison of optimal enforcement levels can therefore be determined by comparing these first-order conditions.

Given that, in the case of a "low-cost" fishery, both regulators' objective functions, (4-5-2) and (4-6-4), are concave, taking the NPIT-optimal level of enforcement $L_e^* > 0$ as that level which in (4-5-5) ensuresthat $\partial V_{R_e}/\partial L_e^* = 0$, and substituting in the CT regulator's first-order condition (4-6-6) reveals that $\tilde{L}_e \gtrless L_e^*$ if and only if

$$(4\text{-}7\text{-}1) \quad [h_1(Q(L_e^*,R,Z),Y)Q(L_e^*,R,Z)+h(Q(L_e^*,R,Z),Y) - C_1(Q(L_e^*,R,Z),\phi)]\frac{\partial Q}{\partial L_e} - b\omega'(L_e) - \Big[[h(Q(L_e^*,R,Z),Y)) - C_1(Q(L_e^*,R,Z),\phi)]\frac{\partial Q}{\partial L_e} - \omega'(L_e)\Big] \gtrless 0$$

Rearranging and simplifying (4-7-1) reveals the condition that $\tilde{L}_e \gtrless L_e^*$ if and only if

$$(4\text{-}7\text{-}2) \qquad h_1(Q(L_e^*,R,Z),Y) \lesseqgtr - \frac{(1-b)\omega'(L_e)}{Q(.)\partial Q/\partial L_e}$$

From Proposition 4-4-1, $\partial Q/\partial L_e < 0$ and, from (4-6-1), $(1 - b)\omega'(L_e) > 0$. The right-hand side of (4-7-2) is then some positive number. Given the assumption that the demand curve is negatively sloped, $h_1(.) < 0$ and hence, from (4-7-2), $\tilde{L}_e > L_e^*$. As $\partial Q/\partial L_e < 0$, this implies that the CT-optimal regulated equilibrium output level is less than that which is optimal for the NPIT regulator.

□

This result is to be expected. The potential gains to producers themselves of controlling the commons exceed those to society as a whole. Once again the CT-optimal regulated equilibrium is equivalent to the optimal behaviour of a cartel that attracts partial funding for its operations externally. The result here, however, as was the case with its counterpart in Chapter Three, suffers from the weakness that to infer the objective of the regulator from the observed enforcement activity requires enough knowledge to solve the optimization problem of each regulator. This difficulty is also associated with responses to changes in the parameter $\phi$ which affects marginal production costs.

Other potential behavioural differences arise in the level and distribution of the output quota, in the response of the regulators to changes in enforcement funding requirements, and potentially in response to a change in the level of consumer income.

Following Proposition 4-5-1, the NPIT regulator is indifferent between any output quota that does not exceed the regulated equilibrium output level consistent with the optimal enforcement level $L_e^*$. From Proposition 4-6-1, however, the CT regulator ensures that the quota is allocated entirely to members of the industry and is set at the regulated equilibrium output level consistent with the CT-optimal enforcement level $\tilde{L}_e$. This ensures that no potential profits are lost through fines incurred on illegal behaviour or through forgone quota rentals.

Secondly, a change in the funding requirements of the resource cost of enforcement does not affect the behaviour of a NPIT regulator. This is because to the NPIT regulator, only interested in aggregate outcome, these funding requirements are transfer payments which have distributional effects but no aggregate impacts. To the CT regulator, however, a change in funding mechanisms represents a change in the cost of gaining profits through regulating itself. An increase in the amount of this cost that the industry must pay reduces the marginal profitability of enforcement at any non-zero enforcement level and thus, as shown in Proposition 4-6-4, reduces the CT-optimal enforcement level.

These results are summarized in the following Proposition.

PROPOSITION 4-7-2: Suppose that an open-access fishery which faces a negatively sloped demand curve is subject to a binding output quota which can be instantaneously and costlessly adjusted by the regulator and which is enforced by means of a constant flat-rate per unit expected penalty that is binding on behaviour. Assuming that the unregulated open-access equilibrium occurs on the upward-sloping portion of the industry supply curve, that enforcing the regulation necessitates the use of scarce resources and incurs a non-decreasing marginal resource cost, and that enforcement is optimal for both regulators, the following results occur:

(i) A CT regulator sets the output quota at the regulated equilibrium output level consistent with the CT-optimal enforcement level and ensures that the quota is allocated entirely to members of the industry. Under NPIT regulation no particular quota size or pattern of distribution can be expected.

(ii) <u>A NPIT regulator is unaffected by changes in the funding requirements for the resource cost of enforcement. A change in the proportion of the resource cost of enforcement funded by the industry itself alters the CT-optimal enforcement level</u>.

The other postulated difference in regulatory behaviour occurs in response to parameter changes. Propositions 4-5-2 and 4-6-2 examined the effect of a change in the level of aggregate consumer income on the behaviour of a NPIT and CT regulator respectively. As shown in the Proof of Proposition 4-7-1 a comparison of the optimal enforcement levels for the two types of regulator can be determined on the basis of a comparison of the regulators' respective first-order conditions with respect to enforcement. So too, following Propositions 4-5-2 and 4-6-2, can a comparison of the response in optimal enforcement levels to a change in the level of aggregate consumer income.

From (4-5-11) and (4-5-12)

$$(4\text{-}7\text{-}3)\quad \frac{dL_e^*}{dY} \gtrless 0 \text{ if and only if } h_2(Q(.),Y) \lessgtr -[h_1(Q(.),Y) - C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial Y}$$

while from (4-6-10), (4-6-14) and (4-6-16)

$$(4\text{-}7\text{-}4)\quad \frac{d\tilde{L}_e}{dY} \gtrless 0 \text{ if and only if } h_2(Q(.),Y) + h_{12}(Q(.),Y)Q(.) \lessgtr -[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y) - C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial Y}$$

where $dL_e^*/dY$ and $d\tilde{L}_e/dY$ denote the effect of a change in aggregate consumer income on optimal enforcement under NPIT and CT regulation respectively. While it is evident, from conditions (4-7-3) and (4-7-4), that the behaviour responses of the two types of regulators will differ, as discussed in Section 3-7, a valid comparison requires enough information to solve the respective optimization problems of the regulator. Here, as there, the complicated nature of behavioural responses is such as to preclude empirical inference and hence the analysis will not be further extended.

Differences in optimal behaviour also arise in the regulation of a "high-cost" fishery.

PROPOSITION 4-7-3: Suppose that an open-access fishery which faces a negatively sloped demand curve is subject to a binding output quota which can be instantaneously and costlessly adjusted by the regulator and which is enforced by means of a constant flat-rate per unit expected penalty that is binding on behaviour. Suppose also that the unregulated open-access equilibrium occurs on the negatively sloped portion of the industry supply curve, that enforcing the regulation necessitates the use of scarce resources and incurs a non-decreasing marginal resource cost but is optimal for both regulators, and that the objective functions of both regulators are concave in enforcement;

(i) Assuming that optimal regulated equilibrium under NPIT and CT hypotheses both occur on the negatively sloped portion of the penalty-inclusive industry supply curve, the NPIT-optimal level of enforcement ($L_e^*$) exceeds (is less than) the CT-optimal level ($\tilde{L}_e$), and hence the NPIT-optimal regulated equilibrium output level is greater (smaller) than that which is optimal for the CT regulator, if and only if the enforcement-induced change in the market value of the existing output level is less than (greater than) the difference between the marginal cost of enforcement to the CT and NPIT regulators. If the marginal cost of enforcement to the CT regulator exceeds that to the NPIT regulator, NPIT-optimal enforcement, and hence equilibrium output level, unambiguously exceeds that which is optimal for the captured regulator.

(ii) Assuming that the optimal regulated equilibria under NPIT and

CT hypothesis both occur on the upward-sloping portion of the penalty-inclusive supply curve but at output levels which exceed the unregulated open-access equilibrium output level, the NPIT-optimal level of enforcement ($L_e^*$) exceeds (is less than) the CT-optimal level ($\tilde{L}_e$) according to the same condition as in part (i) of the Proposition. If the marginal cost of enforcement to the CT regulator is less than that to the NPIT regulator, CT-optimal enforcement exceeds the NPIT-optimal level and hence the CT-optimal regulated equilibrium output level is lower than that which is optimal for the NPIT regulator. This is always the case if the optimal regulated equilibria of both regulators occur on the upward sloping portion of the penalty-inclusive industry supply curve at output levels which are less than that at the unregulated open-access equilibrium level.

Proof.

(i) Following Proposition 4-4-2 the regulated equilibrium in a "high-cost" fishery is unaffected by changes in quota levels such that the quota does not exceed the regulated equilibrium output level consistent with the amount of enforcement. From Proposition 4-5-4 any quota which does not exceed the regulated equilibrium output level consistent with the NPIT-optimal amount of enforcement is optimal for the NPIT regulator while from Proposition 4-6-5 at any regulated equilibrium that occurs at an output level which exceeds the unregulated open-access equilibrium output level the CT-optimal quota is set at the unregulated open-access equilibrium output level. As in Proposition 4-7-1 for the case of a "low-cost" fishery therefore a comparison of optimal enforcement levels can be determined by comparing the appropriate first-order conditions with respect to enforcement. Given the assumption that the objective functions (4-5-2) and (4-6-5) of both regulators are concave in enforcement, taking the NPIT-optimal level of enforcement $L_e^* > 0$ as that level which, in (4-5-5), ensures that $\partial V_{R_e} / \partial L_e^* = 0$ and substituting in the CT regulator's first-order condition (4-6-6), using (4-6-25), reveals that $\tilde{L}_e \gtrless L_e^*$ if and only if

(4-7-5)
$$[h_1(Q(L_e^*,R,Z),Y)Q(L_e^*,R,Z)+h(Q(L_e^*,R,Z),Y) - C_1(Q(L_e^*,R,Z),\phi)]\frac{\partial Q}{\partial L_e} - \Big[b\omega'(L_e)+(1-a)[\lambda(f,L_e)\frac{\partial Q}{\partial L_e} + [Q(L_e^*,R,Z)-R]\frac{\partial \lambda}{\partial L_e}]\Big] - \Big[[h(Q(L_e^*,R,Z),Y) - C_1(Q(L_e^*,R,Z),\phi)]\frac{\partial Q}{\partial L_e} - \omega'(L_e)\Big] \gtrless 0$$

Rearranging and simplifying (4-7-5) reveals the condition that $\tilde{L}_e \gtrless L_e^*$ if and only if

$$(4\text{-}7\text{-}6) \quad h_1(Q(.),Y)Q(.)\frac{\partial Q}{\partial L_e} \gtrless \left[(1-a)[\lambda(f,L_e)\frac{\partial Q}{\partial L_e} + [Q(.)-R]\frac{\partial \lambda}{\partial L_e}] + b\omega'(L_e)\right] - \omega'(L_e)$$

Interpreting (4-7-6), the term $h_1(Q(.),Y)\frac{\partial Q}{\partial L_e}$ represents the enforcement-induced change in demand price. Multiplied by the original regulated equilibrium quantity $Q(.)$, this gives the enforcement-induced change in the market value of the existing level of output. Following (4-6-25), the composite bracketed term on the right-hand side of (4-7-6) shows the marginal cost of enforcement to the CT regulator when operating on the backward-bending portion of the supply curve in a "high-cost" fishery while, from (4-5-1) and (4-5-5), $\omega'(L_e)$ is the marginal cost of enforcement to the NPIT regulator. Given these interpretations of the component terms of (4-7-6) and that, as shown in Proposition 4-4-2, $\partial Q/\partial L_e > 0$ in a "high-cost" open-access fishery, the initial result of part (i) of this Proposition clearly follows.

Following the assumption that the market demand curve is negatively sloped, $h_1(Q(.),Y) < 0$ and hence the left-hand side of (4-7-6) is negative. If the marginal cost of enforcement to the CT regulator exceeds that to the NPIT regulator, the right-hand side of (4-7-6) is strictly positive. Combining these two conditions the negative inequality holds in (4-7-6) implying that the NPIT-optimal level of enforcement is excessive for the CT regulator and hence, by Proposition 4-4-2, that the NPIT-optimal regulated equilibrium output level exceeds that which is optimal for the captured regulator.

(ii) In the case where the optimal regulated equilibria under both hypotheses occur on the upward-sloping portion of the penalty-inclusive industry supply curve at output levels which exceed the unregulated open-access equilibrium output level the condition in part (i) of the Proposition as stated in expression (4-7-6) continues to hold. Here, however, following Proposition 4-4-1, increases in enforcement reduce the regulated equilibrium output level and $\partial Q/\partial L_e < 0$. The left-hand side of (4-7-6) is therefore positive. If then the marginal cost of enforcement to the CT regulator is less than that to the NPIT regulator the right-hand side of (4-7-6) is negative. Combining these two results, the "greater than" inequality holds implying that CT-optimal enforcement exceeds NPIT-optimal enforcement and hence that the CT-optimal regulated equilibrium output level is less than that which is optimal for the NPIT regulator.

In the case where the optimal regulated equilibria under both hypotheses occur on the upward-sloping portion of the penalty-inclusive industry supply curve at output levels less than the unregulated open-access equilibrium output level, it is optimal, following Proposition 4-6-5, for a captured regulator to set the quota equal to the regulated equilibrium output level consistent with the CT-optimal amount of enforcement. Following Proposition 4-5-4 any quota which does not exceed the regulated equilibrium output level consistent with the NPIT-optimal amount of enforcement is optimal for the NPIT regulator. As in Proposition 4-7-1 a comparison of optimal enforcement levels can be determined by comparing the appropriate first-order conditions with respect to employment. In these circumstances the marginal cost of enforcement to the CT regulator consists solely of the proportion of the marginal resource cost of enforcement that the industry is required to fund. Condition (4-7-6) then collapses to that expressed in (4-7-1) and the proof of the result follows as for Proposition 4-7-1.

□

The latter part of result (ii) of the above Proposition is essentially identical to Proposition 4-7-1 as are its interpretation and implications. As explained there, observational inference as to the objectives of the regulator on the basis of these criteria requires enough information to solve the respective optimizing problems of each regulator and hence is not of much empirical use. Result (i) and the remainder of result (ii) suffer from similar limitations. Following Chapter Three, however, there are more discernable behavioural differences. For instance, a regulated equilibrium on the upward-sloping portion of the penalty-inclusive industry supply curve at some output level less than the unregulated open-access equilibrium output level at which non-zero penalties were incurred by the regulated industry would imply the existence of a NPIT regulator. Regrettably though, the non-existence of penalty payments would not unambiguously indicate the presence of a captured regulator.

There is, however, one further implication that can be drawn from Proposition 4-7-3. The per unit marginal benefit of enforcement to the regulator is the difference between marginal revenue and marginal cost which, at regulated equilibria on the backward-bending portion of the penalty-inclusive industry supply curve where marginal cost is negative, is less than that to the NPIT regulator. To complicate matters the marginal resource cost of enforcement to the CT-regulator is also less than that to its NPIT counterpart. The relative magnitudes of optimal enforcement levels then result from a comparison of the marginal penalty cost of enforcement for the CT regulator and the difference in the net, resource-cost inclusive, marginal benefit of enforcement between the two regulators. It is possible that the size of the marginal penalty cost of enforcement to the captured regulator could act to preclude CT regulation in circumstances when it would be optimal for a NPIT regulator to enforce a regulated equilibrium on the backward-bending portion of the penalty-inclusive industry supply curve at some output level higher than that at the unregulated open-access equilibrium level. Alternatively, at regulated equilibria which occur on the upward-sloping section of the penalty-inclusive industry supply curve at output levels smaller than the unregulated open-access equilibrium output level, the net marginal benefit of enforcement to the CT regulator is greater than that to the NPIT regulator. If, therefore, it is optimal for a NPIT regulator to control the fishery to the extent that the regulated equilibrium occurs on the upward-sloping portion of the penalty-inclusive industry supply curve at some output level smaller than the unregulated open-access equilibrium output level, it is profitable for a CT regulator to expand upon the NPIT-optimal enforcement effort.

Together, these two observations imply that, in certain circumstances, regulatory control which results in a regulated equilibrium on the backward-bending portion of the penalty-inclusive industry supply curve is inspired by NPIT objectives and that a zero-penalty regulated equilibrium on the upward-sloping portion of the penalty-inclusive industry supply curve most likely represents the work of a captured regulator. Again the effectiveness of observational inference concerning the objectives of the regulator is somewhat limited in the absence of a priori market information. The informational requirements here, however, are not as comprehensive as those associated with evaluating regulators' behavioural responses to certain parameter changes or with determining objectives on the basis of comparisons of relative enforcement efforts.

## 4-8 CONCLUSION

This chapter applied the regulatory analysis of Chapters Two and Three to the example of an open-access fishery. It began by briefly reviewing the literature on fishery economics. This review followed the steady-state approach which stemmed from the seminal work of Gordon [1954]. The most useful reformulation of Gordon's analysis, for the purposes of this chapter, was that of Copes

[1970, 1972] who combined factor cost and population stock/yield relationships to produce a long run steady-state industry supply curve for the open-access fishery that is backward-bending over a certain price range. While retaining the central theoretical results of other formulations, this allowed the problem to be cast in familiar supply/demand terms.

The main characteristic of an open-access fishery was shown to be that the unregulated open-access equilibrium is economically inefficient in the size of harvest and/or in the amount of factor input. This is analogous to the problem of the tragedy of the commons and results from a lack of well-defined property rights which means that an individual cannot fully appropriate any benefits that accrue from voluntary self-limitation of effort. The inefficiency of the unregulated open-access outcome provides the scope for regulatory control of the fishery.

The model of regulation employed in this chapter was more specialized than that of the preceding two chapters. Regulatory control was exercised through an output quota enforced by means of a flat rate per unit expected penalty which was taken to be an increasing function of the level of resources devoted to enforcement activities. This combination was chosen because, in Chapter Three, it was the combination that generated the clearest contrast between regulatory policy under NPIT and CT hypotheses.

The analysis showed that any regulated equilibrium output between zero and the maximum sustainable yield inclusive could be achieved through the use of an appropriate mixture of quota and expected penalty. A distinction was drawn between "low-cost" and "high-cost" fisheries. As defined in Section 4-4, a "low-cost" fishery is characterized by an unregulated open-access equilibrium which occurs on the upward-sloping portion of the industry supply curve. This equilibrium, although economically inefficient in terms of harvest size, was shown to be technically efficient in factor input. A "high-cost" fishery, by contrast, was defined as one in which the unregulated open-access equilibrium occurs on the backward-bending portion of the supply curve. In this case the initial equilibrium is not only possibly sub-optimal in terms of harvest size but is also technically inefficient in that the same output level could be harvested with a smaller factor input. In a "low-cost" fishery subject to a binding output quota, it was determined that an increase in enforcement reduced the regulated equilibrium output level. In the "high-cost" fishery, however, it was demonstrated that initial increases in enforcement of some binding output quota increase the regulated equilibrium output level. This process continues until the maximum sustainable yield output level is reached beyond which further increases in enforcement act to reduce regulated equilibrium output.

Many of the results derived in this chapter are analogous to those of Chapters Two and Three although in some cases the ambiguities present there were avoided here by the use of the more specialized functional form for the expected penalty. The first of these was that, as in Chapter Three, because of the cost of enforcement, it does not pay a NPIT or CT regulator to respectively restrict output to the socially optimal or pure monopoly profit-maximizing output level.

In the case of a "low-cost" fishery it was shown to be optimal for a CT regulator to restrict the output of the fishery to a greater extent than would a NPIT regulator. This is because competitive behaviour on the commons causes injury to those directly involved in the activity more than to society as a whole. Consequently the potential for appropriation by a CT regulator of the private benefits from restricting activity in the fishery is greater than that of a NPIT regulator which aims to maximize the welfare of society as a whole. This contrasts with the situation in Chapters Two and Three where, in the regulation of an activity which generates a negative externality, the potential benefit for a NPIT regulator from controlling industry output exceeded that from cartelization by a CT regulator.

The other results in the "low-cost" case mirrored those of Chapter Three. Firstly, while some degree of ambiguity may have been reduced here by the use of the flat rate per unit expected penalty, the objective of the regulator could not be inferred from the response of regulatory policy to a change in aggregate consumer income without the possession of enough a priori information to solve the optimization problem of each regulator. Secondly, it was again shown that a CT regulator will ensure that the output quota is set equal to the regulated equilibrium output level and is allocated entirely to members of the regulated industry whereas a NPIT regulator is unconcerned about the size and distribution of the quota, the only proviso being that it does not exceed the regulated equilibrium output level consistent with the optimal amount of enforcement. The third result following from the second was that, of the two, only a captured regulator is affected in its choice of optimal regulatory policy by changes in distributional parameters such as the proportion of the resource cost of enforcement that is funded by the industry.

In the case of a "high-cost" fishery the characteristics of optimal regulatory policy and comparisons of regulatory behaviour depend on whether the regulated equilibria occur on the upward-sloping or backward-bending portions of the penalty-inclusive industry supply curve. It was shown that if the optimum output level of each regulator occurs on the backward-bending portion of the penalty-inclusive industry supply curve, it is profitable for the NPIT regulator to devote a greater amount of resources to enforcement than is optimal for the CT regulator and hence the NPIT-optimal regulated equilibrium output level exceeds the CT-optimal output level. In this circumstance, regulatory objectives could not be readily inferred from the observable policy characteristics of any static regulated equilibrium because all equilibria on the backward-bending portion of the penalty-inclusive supply curve inevitably involved some degree of quota violation and incurred penalty payment. It followed from this, however, that the result concerning the impact of changes in distributional parameters held. As was the case in Chapter Three, a change in regulatory policy in response to a change in the proportion of the resource cost of enforcement that is funded by the industry, or in the percentage of fines that are refunded to members of the industry, reveals the existence of a captured regulator.

It was also shown, in the case of a "high-cost" fishery, that the marginal benefit of enforcement approaches some finite limit value at the regulated equilibrium output level corresponding to the maximum sustainable yield. If it is profitable to do so at this point, further resources will be devoted to enforcement raising the per unit expected penalty thus reducing regulated equilibrium output. In the case where the optimum of each regulator occurs on the upward-sloping portion of the penalty-inclusive supply curve at some output level less than the unregulated open-access equilibrium level, comparisons of regulatory behaviour and inferences about the objectives of the regulator on the basis of such comparisons were shown to proceed as for the regulation of a "low-cost" fishery.

The analysis of this chapter confirmed the central results of Chapters Two and Three, any differences being recognizable as resulting from the different natures of the activities involved. This exercise demonstrated that the model of regulation derived in Chapter Two can be applied across varying types of recognized market failure and that regulatory practices under differently motivated regulators can potentially be readily identified in each situation. Each of the last three chapters, however, has dealt with regulation in a partial equilibrium framework only. The following chapter returns to the regulation of a negative externality, this time produced by a single industry within a simple general equilibrium model of an economy, to determine whether the partial equilibrium results hold also in a general equilibrium context.

NOTES

*An earlier version of this chapter was presented as a paper at the New Zealand Association of Economists' Conference in Wellington, New Zealand, in February 1986. John Yeabsley is thanked for helpful comments.

1. This analysis assumes that the relative slopes of the short-run production functions and the steady-state yield curve $\dot{S} = 0$ are such that each production function intersects the steady-state yield curve only once.
2. The conditions which provide for the instantaneous and continuous emergence of this equilibrium were given in Cheung [1970; 58-62].
3. "Binding on behaviour" in this instance, and elsewhere in the case of a "low-cost" fishery, is defined to mean that the per unit expected penalty does not exceed the difference between the demand price and the penalty-exclusive average cost of production at the quota level R. The level of enforcement which generates an expected penalty rate equal to this difference ensures full compliance with the quota. Following Propositions 2-7-2 and 2-6-1 (iii) further increases in enforcement beyond this full-compliance generating level do not affect the regulated equilibrium output level.
4. The parameter vector Z in (4-4-4) is a composite vector of the parameters which affect either or each of $Q^S(.)$ and $Q^D(.)$ in (4-4-7). The formulation of these functions in (4-4-7) does not include these

parameters. The omission is merely for ease of terminology.

5. In this instance, and elsewhere in the case of a "high-cost" fishery, given that the quota does not exceed the unregulated open-access equilibrium output level while the regulated equilibrium does, any level of enforcement which generates a regulated equilibrium at an output level on the negatively sloped portion of the penalty-inclusive industry supply curve is associated with an expected penalty that is "binding on behaviour". At any level of enforcement greater than that which generates a regulated equilibrium at the MSY output level "binding on behaviour" is defined as in Note 3 for the case of a "low-cost" fishery.

6. The vertical axis intercept of the marginal cost curve at X shows the total cost of catching all the fish in the fishery.

7. Note that as $Y \to \infty$, the NPIT-optimum, where the demand curve and marginal cost curve intersect, tends to the MSY level of output whereas the competitive equilibrium, where the demand curve and average cost curve intersect, tends to zero output. Thus there is always the need for regulatory control of the fishery.

8. Montgomery [1972] shows how this is accomplished through the use of a system of transferable rights. The result is achieved equally well in quantitative terms whether by means of tradeable licences or tradeable output quotas.

9. Following Proposition 4-4-2, an increase in enforcement raises the regulated equilibrium output level of a "high-cost" fishery. As the regulated equilibrium output level approaches the MSY, the two components of the first composite term in (4-5-5), which represents the marginal benefit of enforcement, are affected. The difference between demand price and marginal cost rises but the impact of the marginal unit of enforcement on output falls. The change in the product therefore depends on the relative rates of change of these two components. It is possible that the rates of change may be such that the marginal benefit of enforcement increases more quickly than its marginal cost. In this case, the NPIT regulator's objective function would be non-concave over some output range.

10. This result holds in the case where the quota is constrained, by assumption not to exceed the unregulated competitive equilibrium output level. If no such constraint was imposed, following (4-6-8), the CT-optimal quota would, in the steady state, be set at the regulated equilibrium output level consistent with the CT-optimal amount of enforcement. With a regulated equilibrium output level which exceeded the unregulated competitive equilibrium level this would involve the eventual imposition of a quota at some level higher than the initial output level in the fishery. In order for the interaction between the underlying factor cost and stock/yield relationship necessary to produce the required dynamic

response in the fishery to occur, the quota would initially have to be set at some level below the unregulated competitive equilibrium output level. The "no fines" situation in a "high-cost" fishery with a regulated equilibrium output level which exceeds the unregulated competitive equilibrium level is therefore a steady-state result and penalty payments are incurred by the industry throughout the dynamic adjustment process.

Appendix 4-1.

Following the similar analysis in Chapter Three, several parameters may be assumed to affect the regulated equilibrium. Two such parameters are used here, the level of aggregate income and a technology parameter $\phi$ which reflects the structure of marginal production cost. These parameters affect the demand and supply curves of the fishery respectively. Reformulating (4-4-7) to explicitly include these parameters gives

$$\text{(A4-1-1)} \qquad Q^S(P(f,L_e,Y,\phi),L_e,R,\phi) = Q^D(P(f,L_e,Y,\phi),Y)$$

Totally differentiating (A4-1-1) with $df = dL_e = d\phi = 0$ gives

$$\text{(A4-1-2)} \qquad \frac{\partial Q^S}{\partial P}\frac{\partial P}{\partial Y} - \frac{\partial Q^D}{\partial P}\frac{\partial P}{\partial Y} - \frac{\partial Q^D}{\partial Y} = 0$$

and rearranging

$$\text{(A4-1-3)} \qquad \frac{\partial P^*}{\partial Y} = -\frac{\partial Q^D/\partial Y}{\frac{\partial Q^D}{\partial P} - \frac{\partial Q^S}{\partial P}} \gtrless 0 \text{ if and only if } \frac{\partial Q^D}{\partial Y} \gtrless 0$$

given the assumptions that $\partial Q^D/\partial P < 0$ and $\partial Q^S/\partial P > 0$.

Given the assumption that the output of the fishery is a normal good, $\partial Q^D/\partial Y > 0$, and hence an increase in aggregate income raises the market equilibrium price. Examining the response of supply to the increase in market price

$$\text{(A4-1-4)} \qquad \frac{\partial Q}{\partial Y} = \frac{\partial Q^S}{\partial P}\frac{\partial P^*}{\partial Y} \gtrless 0 \text{ if and only if } \frac{\partial P^*}{\partial Y} \gtrless 0$$

given the assumption that supply increases with price.

Using the result from (A4-1-3) then, an increase in aggregate consumer income increases the regulated equilibrium output level of the industry.

Totally differentiating (A4-1-1) with $df = dL_e = dY = 0$ gives

$$(A4\text{-}1\text{-}5) \qquad \frac{\partial Q^S}{\partial P}\frac{\partial P}{\partial \phi} + \frac{\partial Q^S}{\partial \phi} - \frac{\partial Q^D}{\partial P}\frac{\partial P}{\partial \phi} = 0$$

and rearranging

$$(A4\text{-}1\text{-}6) \quad \frac{\partial P^*}{\partial \phi} = \frac{\partial Q^S/\partial \phi}{\frac{\partial Q^D}{\partial P} - \frac{\partial Q^S}{\partial P}} \gtrless 0 \text{ if and only if } \frac{\partial Q^S}{\partial \phi} \lessgtr 0$$

using assumptions as for (A4-1-3).

Following (4-5-4), an increase in $\phi$ is assumed to raise the marginal cost structure of the fishery at every output level. This causes a reduction in supply at any market price for which supply is defined and hence $\partial Q^S/\partial \phi < 0$. From (A4-1-6) therefore, an increase in $\phi$ increases market equilibrium price which, from the demand curve, is consistent with a reduction in regulated equilibrium quantity and thus $\partial Q/\partial \phi < 0$.

Appendix 4-2.

Given the assumption that (4-6-4) is concave in the level of enforcement, term A, as defined in (4-6-12) is negative.

$$\text{(A4-2-1)}\quad A = [h_1(Q(.),Y)Q(.)+h(Q(.),Y)-C_1(Q(.),\phi)]\frac{\partial^2 Q}{\partial L_e{}^2}$$

$$-\frac{\partial^2 w}{\partial L_e{}^2} + \left[[h_{11}(Q(.),Y)Q(.)+2h_1(Q(.),Y)-C_{11}(Q(.),\phi)]\frac{\partial Q}{\partial L_e}\right]\frac{\partial Q}{\partial L_e}$$

Following Proposition 4-6-1, the initial bracketed term in the right-hand side of (A4-2-1) is negative. It is assumed,as in Chapter 3,that the marginal effectiveness of enforcement as a deterrent diminishes as the level of enforcement rises and thus $\partial^2 Q/\partial L_e{}^2 > 0$. Using (4-6-3) and assuming that the quota is set so that no violations occur, from (4-6-10)

$$\text{(A4-2-2)}\qquad \frac{\partial^2 w}{\partial L_e{}^2} = b\omega''(L_e) > 0$$

given the assumption as in (3-2-2) that the marginal resource cost of enforcement is increasing in enforcement activity. Finally, from Proposition 4-4-1, $\partial Q/\partial L_e < 0$.

Given these results and assumptions on the signs of component terms in (A4-2-1) it is sufficient for $A < 0$ that

$$\text{(A4-2-3)}\quad h_{11}(Q(.),Y)Q(.) + 2h_1(Q(.),Y) - C_{11}(Q(.),\phi) < 0$$

Following the assumptions in (4-5-3) and (4-5-4), $h_1 < 0$ and $C_{11} > 0$. Condition (A4-2-3) is satisfied if and only if;

$$\text{(A4-2-4)}\quad h_{11}(Q(.),Y) < [C_{11}(Q(.),\phi)-2h_1(Q(.),Y)]/Q(.)$$

that is, provided that demand does not become significantly more price responsive as quantity increases.

# CHAPTER FIVE

# REGULATION AND ENFORCEMENT; A SIMPLE GENERAL EQUILIBRIUM ANALYSIS*

## 5-1 INTRODUCTION

Chapters Two, Three, and Four analysed models of regulation and enforcement within a partial equilibrium framework. The results derived were shown to be consistent with those generally found in the deterrence and crime literatures but dependent on the form of expected penalty function used. In addition, the characteristics and properties of the regulated equilibria were shown to differ according to the objectives of the regulator.

The vast majority of the regulatory literature uses partial equilibrium analysis. As argued in Chapter One, Section 1-5, examination of the implications of regulation within a general equilibrium framework, and of the interactions between its costs and benefits when the enforcement process is explicitly modelled in this context, appears to have been neglected.

In light of this, the present chapter extends the analysis of the regulation of an externality from Chapter Two by embedding the regulated industry within a simple general equilibrium model of an economy, the structure of which is outlined in the following section. In each section the results are derived in the general framework and illustrated in the case where preferences are assumed to be Cobb-Douglas.

Following the derivation of a regulated equilibrium in the economy, the properties of regulated equilibria under NPIT and CT regulators are analysed and compared.

## 5-2 COMPETITIVE EQUILIBRIUM WITH AN EXTERNALITY

The economy is assumed to have two final consumption goods and one primary factor of production. $L_i$ denotes the amount of labour employed in industry i, i = 1,2. The production technology is Ricardian with $a_i$ being the input-output coefficient in industry i. This gives production functions,

(5-2-1) $$Q_1 = L_1/a_1$$

(5-2-2) $$Q_2 = L_2/a_2(L_1) \ ; \ a_2(0) = a_2$$

As shown by (5-2-2) the input-output coefficient in industry 2 is a function of the labour employed in industry 1. It is assumed that industry 1 exerts a negative externality on industry 2 so that an increase in employment and output in industry 1 reduces the productivity of labour employed in industry 2. Therefore $a_2'(L_1) > 0$ and the input requirements for a given level of output in industry 2 rise with increased activity in industry 1.

There is a fixed total L of labour available which is assumed to be fully employed between the two industries, thus

(5-2-3) $$L_1 + L_2 = L$$

Substituting for $L_1$ and $L_2$ from (5-2-1) and (5-2-2) gives the equation of the production possibility frontier

(5-2-4) $$a_1Q_1 + a_2(a_1Q_1).Q_2 = L$$

Totally differentiating (5-2-4) gives

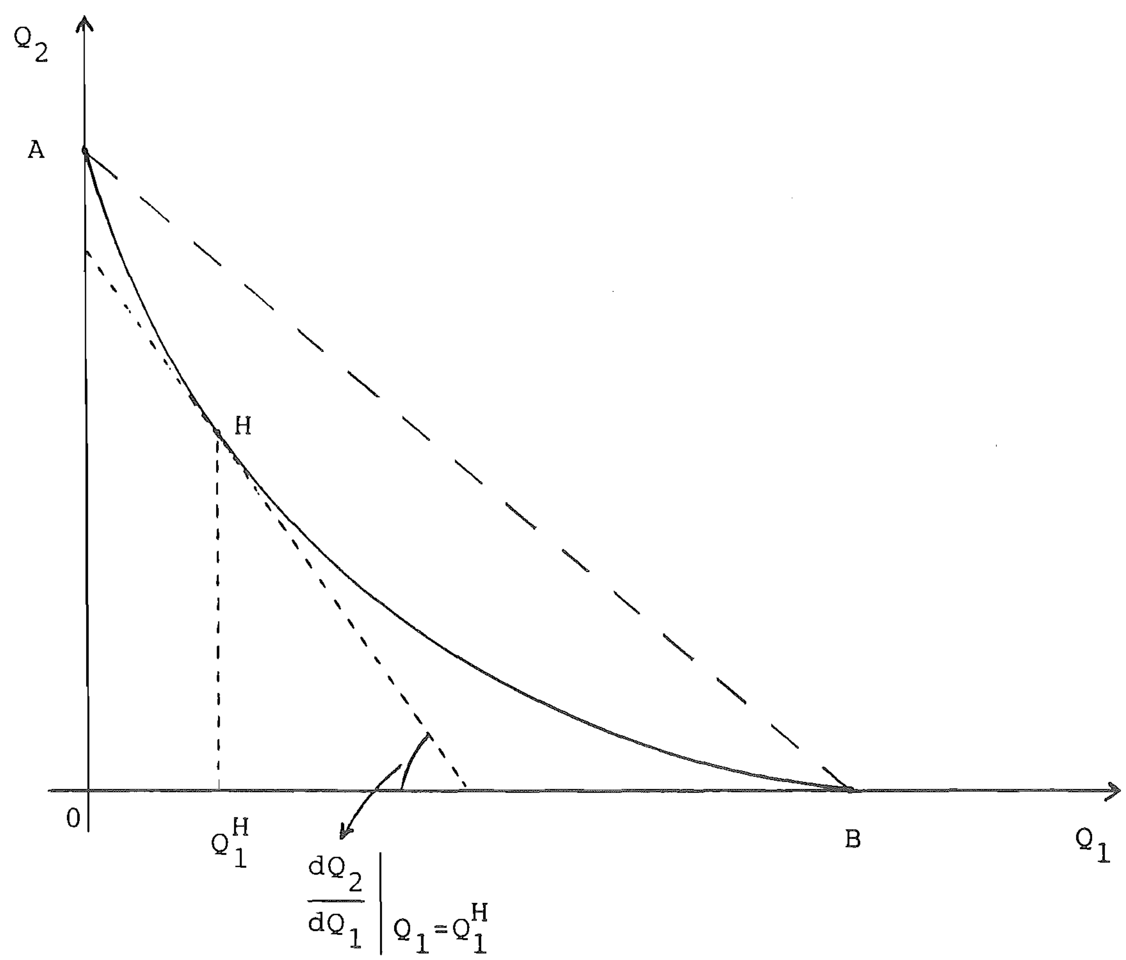
$$a_1 dQ_1 + a_2(a_1Q_1)dQ_2 + a_2'(a_1Q_1)a_1dQ_1 = 0 \tag{5-2-5}$$

and rearranging

$$\frac{dQ_2}{dQ_1} = -\frac{a_1(1+a_2'(a_1Q_1)Q_2)}{a_2(a_1Q_1)} < 0 \tag{5-2-6}$$

Equation (5-2-6) shows the slope of the externality-affected production possibility frontier which is illustrated in Figure 5-2-1 below. The dashed line AB is the usual Ricardian frontier with slope given by the input-output coefficients that would prevail in the absence of any externality. The solid line AHB is the frontier in the presence of the externality.

Figure 5-2-1: Production possibilties in the presence of an externality



From (5-2-1) and (5-2-2), when $L_1 = 0$ no output is produced in industry 1 and hence no externality occurs. Substituting for $a_2(0)$ in (5-2-4) shows that the vertical axis intercept of the externality-affected frontier at A is then identical to that of the usual Ricardian frontier as is the horizontal at B. From (5-2-6) however, the absolute value of the slope given $L_1 = 0$ is greater than in the absence of the externality as $a_2'(0) > 0$.

As shown in (5-2-6) the effect of the externality is such that a unit transfer of labour from industry 2 to industry 1 alters the output of industry 2 not only by the marginal product of labour $a_2(L_1)$ but in addition reduces the productivity of labour remaining in the industry. The slope of the production possibility frontier with the externality is therefore not constant.

In Figure 5-2-1 the externality-affected production possibility frontier is drawn convex to the origin giving a production possibility set that is non-convex. The frontier may be visualized as the outer envelope of Ricardian segments each of which is technically possible only at the tangent point such as H, the slope of the segment being determined by the level of output of industry one at that point. Commencing from point B where $L_1 = L$, successive unit reductions in $L_1$ reduce the output of industry 1 by the constant amount $a_1$. The marginal output from industry 2 however rises as $L_1$ is reduced because the smaller negative externality has the effect of raising the productivity of all labour employed in industry 2. If, as $L_1$ is reduced, the slope of the production possibility frontier is increased, then

$$(5\text{-}2\text{-}7) \qquad \frac{\partial}{\partial Q_1}\left[\frac{dQ_2}{dQ_1}\right] > 0$$

Differentiating (5-2-6) gives

$$(5\text{-}2\text{-}8) \qquad \frac{\partial^2 Q_2}{\partial Q_1^2} =$$

$$-\left[\frac{a_2(a_1Q_1)[a_1^2 a_2''(a_1Q_1)Q_2]-a_1^2[1+a_2'(a_1Q_1)Q_2]a_2'(a_1Q_1)}{[a_2(a_1Q_1)]^2}\right]$$

Expression (5-2-8) is positive if and only if the numerator of the bracketed term is negative. Examination of its component terms shows that a sufficient condition for (5-2-7) to be satisfied is that $a_2''(a_1Q_1) \leq 0$. Rearranging and simplifying reveals the necessary condition

$$(5\text{-}2\text{-}9) \quad a_2''(a_1Q_1) < [1+a_2'(a_1Q_1)Q_2]\frac{a_2'(a_1Q_1)}{a_2(a_1Q_1)Q_2}$$

where the right-hand-side of (5-2-9) is some positive number.

With the presence of an externality, the well known result is that the competitive equilibrium market outcome is not pareto optimal for the economy. Assuming free entry with labour as the numeraire, each industry maximizes profits at the zero level. Price equals average cost in both industries and the relative prices of the two commodities are

$$(5\text{-}2\text{-}10) \qquad \frac{P_1^*}{P_2^*} = \frac{a_1}{a_2(a_1Q_1^*)}$$

where the asterisks denote competitive equilibrium magnitudes.

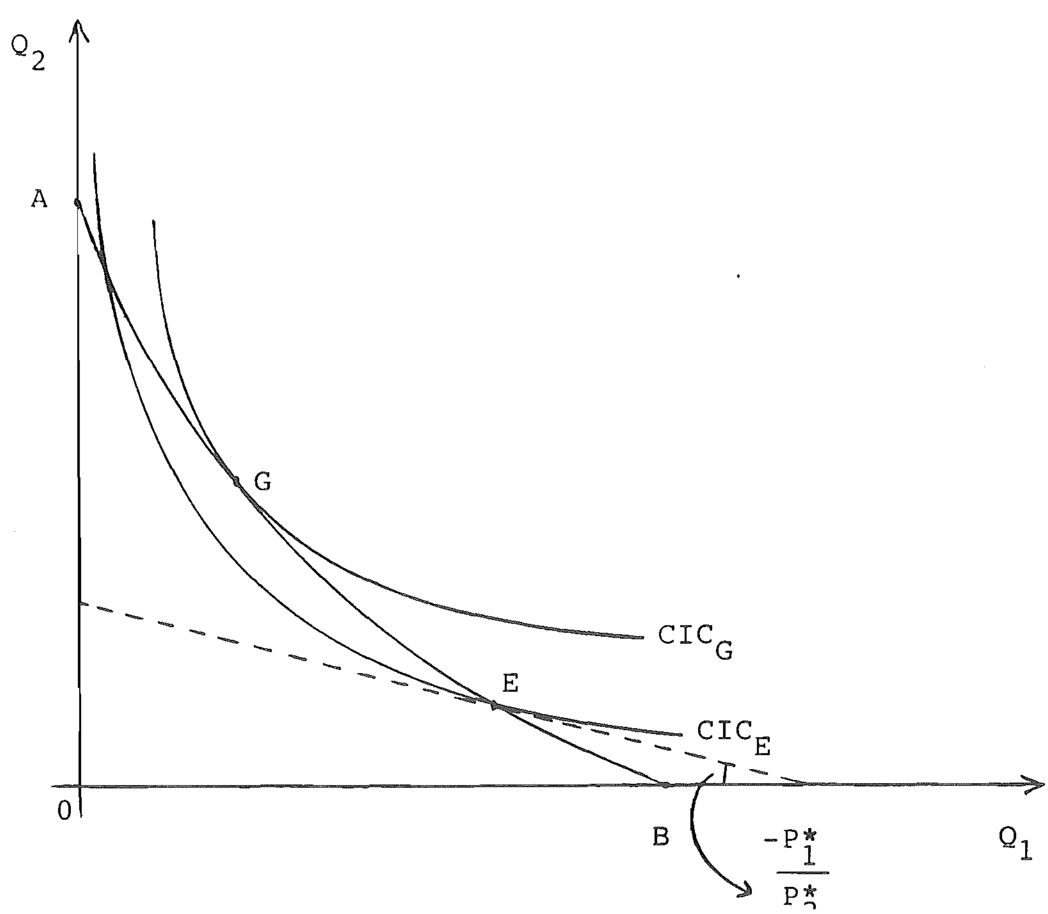
From (5-2-6) and (5-2-10) it is clear that

$$(5\text{-}2\text{-}11) \qquad \frac{dQ_2}{dQ_1} < -\frac{P_1^*}{P_2^*}$$

At the competitive equilibrium the slope of the production possibility frontier is steeper than that of the economy's budget constraint. Assuming that the consumption side of the economy comprises a finite number of individual agents and that income effects are identical over the relevant income range, the preferences of these agents can be represented by community indifference curves. Equation (5-2-11) then implies that, at the competitive equilibrium, the rate of commodity substitution of commodity 1 for commodity 2 is less than the rate of product transformation.

This is illustrated in Figure 5-2-2 below. The competitive equilibrium occurs at point C where a community indifference curve is tangential to a national budget constraint with slope $- P_1^*/P_2^*$.

Figure 5-2-2: Comparison of the competitive equilibrium and pareto optimum



Denoting by $U(Q_1,Q_2)$ the utility function which generates the community indifference curves, the pareto optimum for the economy occurs where the utility function is maximized with respect to the consumption of both commodities subject to the technological and resource constraints of production. Forming the Lagrangean

$$(5\text{-}2\text{-}12)\quad L(Q_1,Q_2,L_1,L_2) \equiv U(Q_1,Q_2) + \lambda_1[a_1Q_1-L_1] + \lambda_2[a_2(a_1Q_1)Q_2-L_2] + \lambda_3[L-L_1-L_2]$$

Assuming an interior optimum which is technically efficient and such that all labour is employed, the first-order conditions are as follows.

$$(5\text{-}2\text{-}13)\quad \frac{\partial L}{\partial Q_1} = \frac{\partial U}{\partial Q_1} + \lambda_1 a_1 + \lambda_2 a_1 a_2'(a_1Q_1)Q_2 = 0$$

$$(5\text{-}2\text{-}14)\quad \frac{\partial L}{\partial Q_2} = \frac{\partial U}{\partial Q_2} + \lambda_2 a_2(a_1Q_1) = 0$$

$$(5\text{-}2\text{-}15)\quad \frac{\partial L}{\partial L_1} = -\lambda_1 - \lambda_3 = 0$$

$$(5\text{-}2\text{-}16)\quad \frac{\partial L}{\partial L_2} = -\lambda_2 - \lambda_3 = 0$$

Elimination of $\lambda_1$, $\lambda_2$ and $\lambda_3$ gives

$$(5\text{-}2\text{-}17)\quad \frac{\partial U}{\partial Q_1} - \frac{a_1 \partial U/\partial Q_2}{a_2(a_1Q_1)} - a_1 a_2'(a_1Q_1)\frac{\partial U/\partial Q_2}{a_2(a_1Q_1)} = 0$$

and rearranging

$$(5\text{-}2\text{-}18)\quad \frac{\partial U(Q_1^{**},Q_2^{**})/\partial Q_1}{\partial U(Q_1^{**},Q_2^{**})/\partial Q_2} = \frac{a_1(1+a_2'(a_1Q_1^{**})Q_2^{**})}{a_2(a_1Q_1^{**})}$$

where the double asterisks denote optimal magnitudes.

That is, the rate of commodity substitution equals the rate of product transformation at the pareto optimum which occurs at point G in Figure 5-2-2. With a diminishing marginal rate of substitution, the decentralized competitive outcome in the presence of an externality results in an excessive output from industry 1 as neither producers nor consumers face prices which accurately reflect the true cost of their actions. As a consequence, the level of social welfare is below that potentially attainable and could be improved if industry 1 was forced to contract provided that any such reduction involved no other effects.

The competitive equilibrium is now illustrated in the case when preferences are represented by the Cobb-Douglas direct utility function

$$(5\text{-}2\text{-}19) \qquad u(q_1,q_2) = q_1^b q_2^{1-b} \quad , \quad 0 < b < 1$$

where $q_i$ represents the quantity consumed of each commodity.

Maximizing (5-2-19) with respect to $q_i$ subject to the national budget constraint gives the demand functions

$$(5\text{-}2\text{-}20) \qquad q_1 = \frac{bL}{P_1} \quad , \quad q_2 = \frac{(1-b)L}{P_2}$$

Using (5-2-1), (5-2-2) and (5-2-10) in (5-2-20), the competitive equilibrium employment levels in each industry are

$$(5\text{-}2\text{-}21) \qquad L_1^* = bL \ , \ L_2^* = (1-b)L$$

Expression (5-2-21) describes the competitive equilibrium in the Cobb-Douglas case. As in the more general case above, the competitive equilibrium is not pareto optimal and hence potential welfare gains exist.

## 5-3 REGULATED EQUILIBRIUM

The simple model of Section 5-2 is now complicated by the introduction of a regulatory agency which seeks to reduce the output of industry 1. As in the partial equilibrium framework of Chapter Two, it is theoretically possible, in the absence of enforcement considerations, to devise a set of Pigouvian taxes and/or subsidies which transform relative prices to those which generate the pareto optimum or, alternatively, by quantitative methods to restrict the output of industry 1 to the optimal level. However, as previously mentioned, this ignores the problem of evasion.

It may be socially optimal to devise such instruments to remove the divergence between social and private cost but it is unlikely to be in the private interest of the offending group. Individual agents optimize with respect to private variables and the competitive outcome reflects agents' profit maximizing decisions. Any policy which seeks to adversely affect this outcome for any individual will therefore be ignored unless there exist either incentives for compliance or deterrents to non-compliance.

Sections 3-2 and 3-4 demonstrate that the most significant differences in regulatory behaviour occur when quantitative techniques are employed. It was shown in Proposition 3-2-1, for the case of an expected penalty function that is strictly convex in output and the extent of constraint violation, that NPIT regulators will not use non-zero quotas but it was also shown that the use of this type of quantitative instrument, albeit at the zero level, is not dominated by any other instrument type under either of the hypotheses on regulatory objectives analysed. For this reason, the regulator in the present context is assumed to set output quotas for firms in industry 1.

In the absence of penalties to deter non-compliance, as mentioned above, individuals' self interest will result in the competitive equilibrium output levels being produced in each industry and the quotas will be ineffective. To facilitate compliance it is assumed that the regulatory agency is able to inflict punishment on offenders against the regulation. A regulated equilibrium occurs when supply and demand are equated for all commodities and firms have maximized their expected profits within the regulated environment.

An industry quota of R units of commodity 1 is imposed by the regulatory authority at a value which is exceeded at the unregulated competitive equilibrium. Thus

$$(5\text{-}3\text{-}1) \qquad 0 \leq R < Q_1^*$$

The quota is enforced by a monetary penalty which is proportional to production in excess of the quota owned by the individual firm. This penalty takes the form of a constant marginal fine of $a_3$ per unit of illegal production. Not all offenses are detected however. The probability of detection, and successful prosecution, of a firm which exceeds its quota is a function $f(L_e)$ where $L_e$ is the amount of labour employed by the regulatory agency. The expected penalty per unit of excess production is then

$$(5\text{-}3\text{-}2) \qquad \rho(a_3, L_e) = a_3 f(L_e) \quad ; \; f'(L_e) > 0 \; , \; 0 \leq f(L_e) \leq 1$$

Firms are assumed to be risk neutral. Their behaviour therefore is motivated by the absolute size of the expected penalty rather than its component parts. As before

(5-3-3) $\rho(0,L_e) = \rho(a_3,0) = 0$

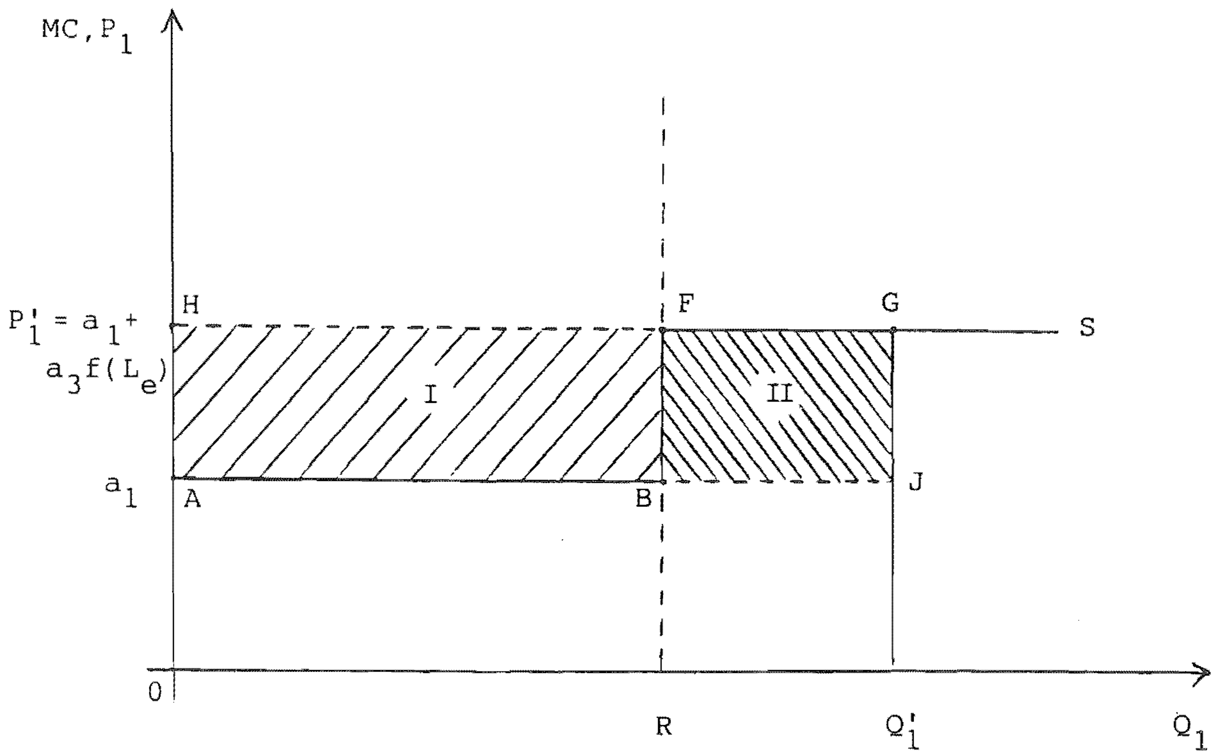
and enforcement requires the employment of scarce, otherwise productive resources.

The regulated equilibrium output of industry 1 is unaffected by the assignment of quota to individual firms. Quotas are tradeable and become an asset which acquires its value from the marginal expected penalty which is incurred when a firm's production exceeds the share of the industry quota that it holds. This becomes important when analyzing the output decisions of firms and for defining income.

With free entry and constant marginal costs the size of the individual firm is indeterminate and irrelevant so that the effects of the quota can be analysed at the industry level. Beyond the quota level R, the cost of production rises by the amount of the expected penalty. This produces the stepwise supply curve ABFG shown in Figure 5-3-1 overleaf.

With free entry, the expected unit fine becomes an additional cost of production whether or not a firm exceeds its holding of quota. This is because, allowing for, and in excess of, the resource cost of production which is unaffected by the regulatory environment, with a tradeable quota, the per unit expected penalty represents the minimum cost of producing an additional illegal unit and the maximum amount paid to acquire a legal right to produce.

Figure 5-3-1: Supply and income in the regulated environment




Thus in equilibrium the price of commodity 1 is

$$P_1 = a_1 + a_3f(L_e) \tag{5-3-4}$$

The quota is valuable because it allows its holder the right to produce a certain amount of the commodity without attracting a penalty. This confers a rent to the initial quota holders of amount $a_3f(L_e)R$ which is shown on Figure 5-3-1 by area I. At any output $Q_1' > R$, which is the equilibrium level consistent with the price $P_1'$, total fine payments of $a_3f(L_e)[Q_1'-R]$ are incurred. These are shown by area II.

It is assumed that any such fines are redistributed to all agents in the economy in lumpsum fashion. The only other source of income in the economy is labour income. Therefore the before tax income of consumers is

$$L + a_3f(L_e)R + a_ef(L_e)[Q_1-R] = L + a_3f(L_e)Q_1 \tag{5-3-5}$$

which is independent of the quota. This is consistent with Sections 3-2 and 3-4 where it was shown that industry equilibrium output given a constant marginal expected penalty is independent of the quota level and its initial allocation. The aggregate size of area AHGJ may then be invariant to changes in the quota level but the relative sizes of areas I and II are not. The quota thus has important distributional effects which will be discussed in Section 5-5.

As in the partial equilibrium analysis, it is assumed that enforcement activities are funded by a lump-sum tax. From (5-3-5) the after-tax income of consumers is[1]

$$y(a_3,L_e) = L + a_3f(L_e)Q_1 - L_e \tag{5-3-6}$$

Free entry ensures that in industry 2 the equilibrium price equals average cost, as without regulation. That is

$$P_2 = a_2(a_1Q_1) \tag{5-3-7}$$

The demand for commodity 1, $D_1(P_1,P_2,y)$ depends on prices and national income within the regulatory environment. Given the level of enforcement $L_e$ and the output of commodity 1, $Q_1$, the demand for that commodity is

$$q_1 = D_1(a_1+a_3f(L_e),\, a_2(a_1Q_1),\, L + a_3f(L_e)Q_1 - L_e) \tag{5-3-8}$$

A <u>regulated equilibrium</u> occurs when demand and supply for commodity 1 are equal. Using (5-3-8), this requires that excess demand is zero. That is,

$$E \equiv D_1(a_1+a_3f(L_e),\, a_2(a_1Q_1),\, L + a_3f(L_e)Q_1 - L_e) - Q_1 = 0 \tag{5-3-9}$$

Labour market equilibrium is contained within the budget constraint and Walras law ensures that the market for commodity 2 is in equilibrium when (5-3-9) holds. It is assumed that excess demand for commodity 1 falls as the supply of output from industry 1 rises, thus

$$(5\text{-}3\text{-}10) \qquad \frac{\partial E}{\partial Q_1} = \frac{\partial D_1}{\partial P_2} \cdot a_2'(a_1Q_1)a_1 + \frac{\partial D_1}{\partial y} \cdot a_3 f(L_e) - 1 < 0$$

This guarantees a unique solution to (5-3-9), which is independent of the quota R. This condition also guarantees that the regulated equilibrium output of commodity 1 can be expressed as a function of the effort devoted to enforcing the quotas and the per unit fine, $Q_1(a_3, L_e)$. From (5-3-9)

$$(5\text{-}3\text{-}11) \qquad \frac{\partial Q_1}{\partial a_3} = -\left(\frac{\partial D_1}{\partial P_1} f(L_e) + \frac{\partial D_1}{\partial y} f(L_e)Q_1\right)/\frac{\partial E}{\partial Q_1}$$

Slutsky's equation simplifies the numerator, so that

$$(5\text{-}3\text{-}12) \qquad \frac{\partial Q_1}{\partial a_3} = -f(L_e)\left.\frac{\partial D_1}{\partial P_1}\right|_u / \frac{\partial E}{\partial Q_1} < 0$$

That is, an increased penalty for exceeding the quota reduces the regulated equilibrium output of commodity 1. However,

$$(5\text{-}3\text{-}13) \qquad \frac{\partial Q_1}{\partial L_e} = -\left(\frac{\partial D_1}{\partial P_1} a_3 f'(L_e) + \frac{\partial D_1}{\partial y}[a_3 f'(L_e)Q_1 - 1]\right)/\frac{\partial E}{\partial Q_1}$$

Applying Slutsky's equation to the numerator of (5-3-13) gives

$$(5\text{-}3\text{-}14) \qquad \frac{\partial Q_1}{\partial L_e} = -\left(a_3 f'(L_e)\left.\frac{\partial D_1}{\partial P_1}\right|_u - \frac{\partial D_1}{\partial y}\right)/\frac{\partial E}{\partial Q_1}$$

Hence

$$(5\text{-}3\text{-}15) \qquad \frac{\partial Q_1}{\partial L_e} \lesseqgtr 0 \text{ if and only if } a_3 f'(L_e)\left.\frac{\partial D_1}{\partial P_1}\right|_u - \frac{\partial D_1}{\partial y} \lesseqgtr 0$$

Examination of (5-3-15) shows that a necessary condition for an increase in enforcement to increase the output of industry 1 is that commodity 1 is an inferior good. It is sufficient however for $\partial Q_1/\partial L_e < 0$ that commodity 1 is not inferior. The analysis here uses community indifference curves to represent aggregate welfare. This requires the global aggregation of individual preferences. For global aggregation of individual preferences, it is necessary that the commodities exhibit unitary income elasticities [Manning, 1986]. Commodity 1 therefore cannot be inferior and thus sufficiency for $\partial Q_1/\partial L_e < 0$ is assured.

Further examination of (5-3-9) reveals that the regulated equilibrium output of industry 1 is independent of the quota level R. This follows from the previous discussion of (5-3-6). With a constant marginal expected penalty the introduction of a binding quota has a once and for all effect in the model assuming it is enforced. The character of the economy is transformed from individual competitive optimization into constrained maximization within a regulatory framework. The one aggregate function of the quota level is to set a lower bound on the degree to which the economy can be controlled. Thus

$$(5\text{-}3\text{-}16) \qquad R \leq Q_1(a_3, L_e)$$

and enforcement determines the range in which the quota applies. In the justification of the underlying assumptions of this model it was made clear that a binding regulatory instrument impinges on individual plans and decisions, and must be enforced to secure some degree of effective control. No profit maximizing firm within a competitive environment therefore will operate at a lower level than their holding of quota allows.

Using the assumption from (5-3-3) in (5-3-9) implies that

$$(5\text{-}3\text{-}17) \qquad Q_1(a_3,0) = Q_1(0,L_e) = Q_1^*$$

so that, here, as in the partial equilibrium framework of Chapter Three, with no enforcement the regulated equilibrium coincides with the unregulated competitive equilibrium.

It is assumed that initial enforcement efforts raise the probability of detection quickly thus

$$(5\text{-}3\text{-}18) \qquad \lim_{L_e \to 0} f'(L_e) = \infty$$

Using this assumption in (5-3-14) gives

$$(5\text{-}3\text{-}19) \qquad \left.\frac{\partial Q_1}{\partial L_e}\right|_{L_e=0} < 0$$

so that initial enforcement efforts will reduce the regulated equilibrium output of commodity 1.

The function $Q_1(a_3,L_e)$ allows the remaining economic variables also to be expressed as functions of the unit fine and labour devoted to enforcement.

The modified form of the labour constraint from (5-2-3) is

$$(5\text{-}3\text{-}20) \qquad L_1 + L_2 + L_e = L$$

Using this in (5-2-2), substituting for $L_1$ from (5-2-1), and for $Q_1$ from above, gives the regulated equilibrium output of commodity 2.

$$(5\text{-}3\text{-}21) \qquad Q_2(a_3,L_e) = \frac{L-L_e-a_1Q_1(a_3,L_e)}{a_2(a_1Q_1(a_3,L_e))}$$

Substituting the expression for regulated equilibrium output of industry 1 into (5-3-7) gives

$$(5\text{-}3\text{-}22) \qquad P_2(a_3,L_e) = a_2(a_1(Q_1(a_3,L_e)))$$

and into (5-3-6) gives

$$(5\text{-}3\text{-}23) \qquad y(a_3,L_e) = L + a_3f(L_e)Q_1(a_3,L_e) - L_e$$

The regulated equilibrium price of commodity 2, and the after-tax income, are expressed as functions of the unit fine and the labour devoted to enforcement by (5-3-22) and (5-3-23) respectively.

Differentiating (5-3-21) with respect to the unit fine and using (5-3-12) gives

$$(5\text{-}3\text{-}24) \qquad \frac{\partial Q_2}{\partial a_3} = $$

$$\frac{-a_1\frac{\partial Q_1}{\partial a_3}[a_2(a_1Q_1(a_3,L_e))+a_2'(.)(L-L_e-a_1Q_1(a_3,L_e))]}{[a_2(a_1Q_1(a_3,L_e))]^2} > 0$$

hence an increase in the unit fine will increase the output of the commodity adversely affected by the externality.

Differentiating (5-3-21) with respect to enforcement activity gives

$$(5\text{-}3\text{-}25) \qquad \frac{\partial Q_2}{\partial L_e} = -\frac{(1+a_1\frac{\partial Q_1}{\partial L_e})a_2(.)+(L-L_e-a_1Q_1(a_3,L_e))a_2'(.)a_1\frac{\partial Q_1}{\partial L_e}}{[a_2(a_1Q_2(a_3,L_e))]^2}$$

which implies that

$$(5\text{-}3\text{-}26) \qquad \frac{\partial Q_2}{\partial L_e} < 0 \quad \text{if} \quad \frac{\partial Q_1}{\partial L_e} \geq 0$$

while

$$(5\text{-}3\text{-}27) \qquad \frac{\partial Q_2}{\partial L_e} > 0 \text{ only if } \frac{\partial Q_1}{\partial L_e} < 0$$

Interpretation of (5-3-25) shows that additional enforcement of the regulations might alter the output of this commodity in either direction although (5-3-26) and (5-3-27) provide clues about this effect.

From (5-3-25) a sufficient condition for enforcement to increase the regulated equilibrium output of commodity 2 is that

$$(5\text{-}3\text{-}28) \qquad \frac{\partial Q_1}{\partial L_e} < -\frac{1}{a_1}$$

Recalling from (5-2-1) that $a_1$ is the input-output coefficient in industry 1, (5-3-28) shows that, if more than one unit of labour is released from industry 1 as a result of an increase in enforcement, the equilibrium output of commodity 2 rises.

This result is intuitively obvious. A unit increase in labour devoted to enforcement reduces the productive labour force by one unit. Any reduction in labour employed in industry 1, and hence in the output of commodity 1, increases the productivity of labour employed in industry 2. If the reduction in labour employed in industry 1 exceeds the increase in labour devoted to enforcement activities, the net result is an increase in labour employed in industry 2 and hence the output of industry 2 must rise.

Rearranging the numerator of (5-3-25) reveals the necessary and sufficient condition for enforcement to increase the regulated equilibrium output of commodity 2.

$\frac{\partial Q_2}{\partial L_e} \gtrless 0$ if and only if

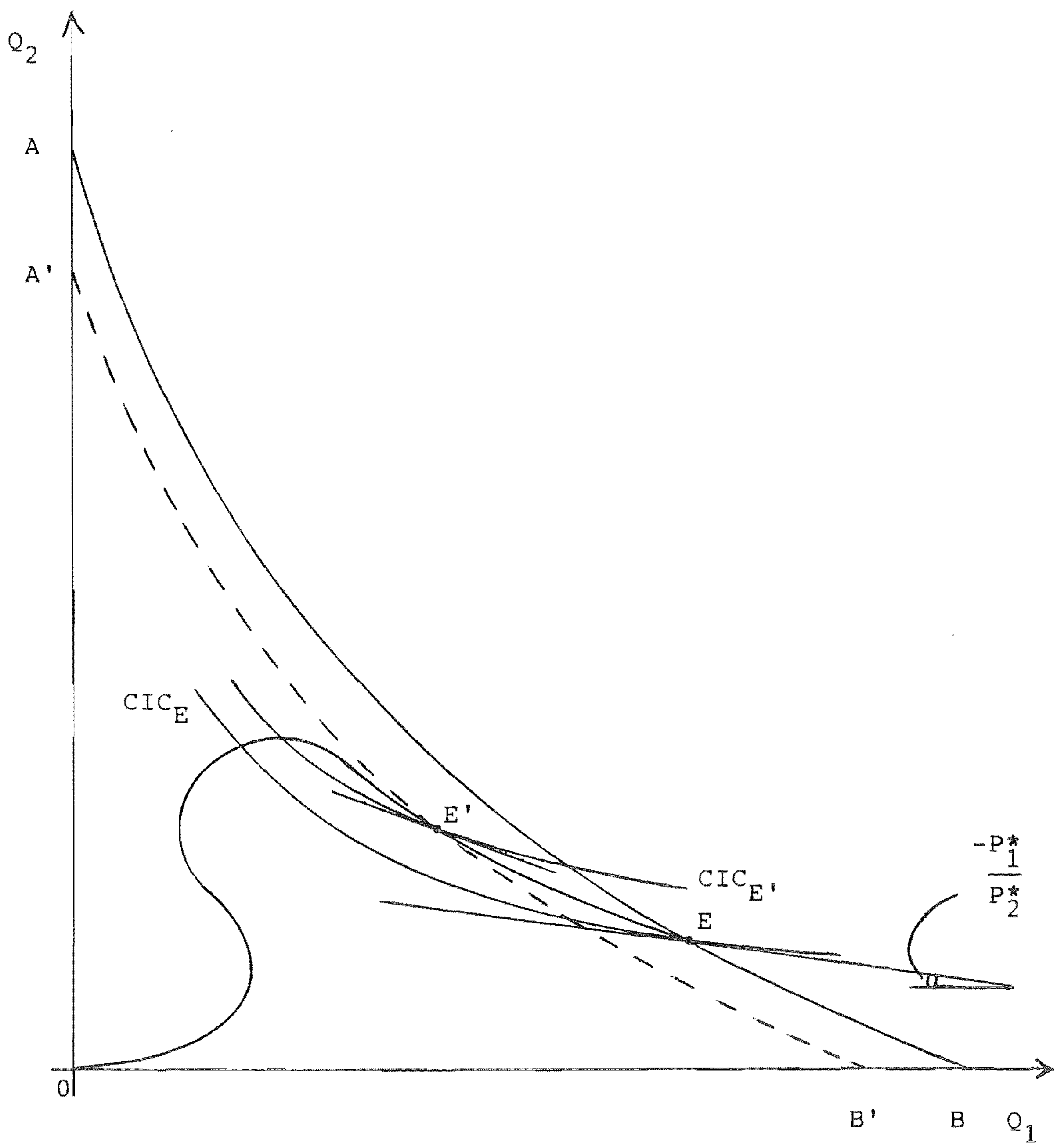
$$(5\text{-}3\text{-}29) \qquad \frac{\partial Q_1}{\partial L_e} \lesseqgtr - \frac{a_2(a_1Q_1(a_3,L_e))}{a_1[a_2(a_1Q_1(a_3,L_e))+a_2'(.)[L-L_e-a_1Q_1(a_3,L_e)]]}$$

Comparing the right-hand sides of (5-3-28) and (5-3-29) shows that it is not necessary for net employment in industry 2 to rise in order for the equilibrium output of commodity 2 to increase with enforcement activity. What is necessary for $\partial Q_2/\partial L_e > 0$ is that the output effect of any net reduction in employment in industry 2, which may occur as a result of increased enforcement activity, is more than offset by the efficiency gains in the production of commodity 2 caused by the fall in the output of commodity 1 which is generated by such activity.

For a given unit fine $a_3$, the functions $Q_1(a_3,L_e)$ and $Q_2(a_3,L_e)$ define a set of regulated equilibria. There is a regulated equilibrium output of each commodity when the input into enforcement is known. These equilibria are illustrated in Figure 5-3-2 overleaf.

As in Figure 5-2-2 the production possibility frontier in the absence of regulation, or enforcement, is AB, and the unregulated competitive equilibrium is at E where a community indifference curve is tangential to a relative price line. If labour is devoted to regulatory enforcement, the productive capacity of the economy is reduced to some level such as A'B'. Enforcement increases the absolute and relative price of commodity 1 and a new equilibrium emerges at a tangency point such as E' which is consistent with preferences and technology. E' is a regulated

Figure 5-3-2: Locus of regulated equilibria




equilibrium. The set of all regulated equilibria is illustrated by the locus OE'E. When no enforcement is attempted the competitive equilibrium at E is achieved. If all labour is devoted to enforcement then nothing can be produced, and the regulated equilibrium is at 0. The shape of the locus is consistent with (5-3-26) and (5-3-27).

It is convenient to summarize the conclusions of this section in the following proposition.

PROPOSITION 5-3-1: If an externality is regulated by an output quota which is enforced by fines, then in a regulated equilibrium, the prices and quantities of commodities depend on the unit fine and the labour devoted to regulatory enforcement but are independent of the size and distribution of the quotas themselves. Increases in the unit fine decrease the output of the externality-generating industry and increase the output of the other commodity. Starting from zero, an increase in enforcement will reduce the output of the externality generating industry. It is possible that the reverse happens as increased enforcement occurs from a higher initial level, but then the output of the other commodity will fall.

The regulated equilibrium is now illustrated for the case when preferences can be represented by the Cobb-Douglas direct utility function given in (5-2-19).

Using the demand function for commodity 1 derived from the first-order conditions of the maximization of (5-2-19) subject to the regulatory budget constraint (5-3-23), (5-3-9) becomes

$$(5\text{-}3\text{-}30) \qquad \frac{b(L+a_3 f(L_e)Q_1-L_e)}{a_1 + a_3 f(L_e)} - Q_1 = 0$$

which implies that

$$(5\text{-}3\text{-}31)\qquad Q_1(a_3, L_e) = \frac{b(L-L_e)}{a_1 + (1-b)a_3 f(L_e)}$$

The effects of changes in the component parts of the expected penalty on the regulated equilibrium can now be examined. From (5-3-31)

$$(5\text{-}3\text{-}32)\qquad \frac{\partial Q_1}{\partial a_3} = \frac{-b(L-L_e)(1-b)f(L_e)}{(a_1 + (1-b)a_3 f(L_e))^2} < 0$$

while

$$(5\text{-}3\text{-}33)\qquad \frac{\partial Q_1}{\partial L_e} = \frac{-b(a_1+(1-b)a_3 f(L_e))-b(L-L_e)(1-b)a_3 f'L_e)}{(a_1 + (1-b)a_3 f(L_e))^2} < 0$$

In the Cobb-Douglas case therefore, an increase in enforcement unambiguously reduces the regulated equilibrium output of commodity 1. For any binding quota within the regulatory framework, an increase in resources devoted to enforcement which increases the expected penalty faced on illegal production will reduce the equilibrium output and employment levels in industry 1 thus lessening the degree of violation of the regulatory constraint and reducing the externality.

This result follows immediately from (5-3-15) noting that with Cobb-Douglas preferences $\partial D_1/\partial y$ is a positive constant. The result is consistent with that widely held in the crime literature where it is claimed that an increase in the expected penalty reduces illegal activity [Becker, 1968; Ehrlich, 1972, 1973; Stigler, 1970]. It represents, within the particular general equilibrium framework used here, an extension of Proposition 2-6-2 where the same result is derived in the partial equilibrium context.

From (5-3-24) and (5-3-32), as $\partial Q_1/\partial a_3 < 0$, an increase in the unit fine raises the regulated equilibrium output of industry 2 while, from (5-3-25), increasing enforcement may expand or contract the regulated equilibrium size of industry 2 depending on the magnitude of $\partial Q_1/\partial L_e$ in (5-3-33).

Equation (5-3-31) also shows that here, as in the general case summarized by Proposition 5-3-1, the regulated equilibrium output of commodity 1 is independent of the size and distribution of the output quota.

These results are summarized by the following Proposition.

PROPOSITION 5-3-2: If an externality is regulated by an output quota which is enforced by fines, and preferences are Cobb-Douglas, then in a regulated equilibrium, the prices and quantities of commodities depend on the unit fine and the labour devoted to regulatory enforcement but are independent of the size and distribution of quotas. Increases in the unit fine decrease the output of the externality-generating industry and increase the output of the other commodity. Increases in enforcement always reduce the output of the externality-generating industry. The output of the other commodity may rise or fall with increases in enforcement activity depending on the magnitude of the enforcement-induced reduction

<u>in the output of the externality-generating industry</u>.

This section has illustrated the mechanisms through which regulation affects the economy and derived the properties of feasible regulated equilibria. In the partial equilibrium analysis of Chapters Three and Four it was shown that the policy mix which the regulatory body employs will depend on its objectives and that it is possible to infer those objectives from the observed combination of instruments used in any particular instance. The following analysis examines the same contrasting regulatory objectives to determine whether the results hold in the context of the particular general equilibrium framework used here.

## 5-4 REGULATION ACCORDING TO NPIT

Assume initially that the regulatory agency is operating according to the Naive Public Interest theory of regulation. The regulatory agency then has as its primary objective the maximization of social welfare. As each agent has an identical utility function, community indifference curves apply irrespective of the level and distribution of after-tax income. Whatever distribution of income is optimal, the aggregate utility function should be maximized. The direct utility function is $u(Q_1, Q_2)$ and the corresponding indirect utility function is $v(p_1, p_2, y)$. Both of these will be used as they provide alternative characterizations of NPIT optimal regulation.

Substituting the regulated equilibrium quantities of commodities 1 and 2 into the direct utility function gives utility as a function of the labour devoted to enforcement and of the unit fine. Recalling the argument of Section 2-7 that it is necessary, in order to maintain marginal deterrents which prevent spillovers between classes of offence, to have a well graduated scale of finite fines [Stigler, 1970], in the present analysis the unit fine will be taken as fixed at its value $a_3$. Therefore

$$(5\text{-}4\text{-}1) \qquad U(L_e) \equiv u(Q_1(a_3,L_e),Q_2(a_3,L_e))$$

This is maximized by the regulator. The first-order condition is

$$(5\text{-}4\text{-}2) \qquad \frac{u_2}{u_1} \leq -\frac{\partial Q_1/\partial L_e}{\partial Q_2/\partial L_e}\ , \ < \text{only if } L_e^* = 0$$

The left-hand side is the rate of commodity substitution in consumption. The right-hand side is the rate of product transformation through regulatory enforcement. Geometrically, the NPIT-optimal regulation, assuming an interior optimum, is found where the set of regulatory equilibria is tangential to a community indifference curve. In Figure 5-3-2 this occurs at a smaller output of commodity 1 than is produced at the illustrated regulated equilibrium E'.

Substituting the demand equations found from equations (5-3-4), (5-3-27) and (5-3-28) into the direct utility function yields the alternative indirect utility function which also relates utility to enforcement.

$$(5\text{-}4\text{-}3) \qquad U(L_e) \equiv V(L_e) = v(P_1(a_3,L_e),\ P_2(a_3,L_e),\ y(a_3,L_e))$$

The first-order condition for a maximum is

$$(5\text{-}4\text{-}4) \qquad \frac{\partial v}{\partial P_1}\cdot\frac{\partial P_1}{\partial L_e}+\frac{\partial v}{\partial P_2}\cdot\frac{\partial P_2}{\partial L_e}+\frac{\partial v}{\partial y}\cdot\frac{\partial y}{\partial L_e} \leq 0\ ;< \text{only if } L_e^* = 0$$

Rearranging (5-4-4) and using Roy's Identity reveals that

$$(5\text{-}4\text{-}5) \quad \frac{\partial v}{\partial L_e} \leq Q_1 \cdot \frac{\partial P_1}{\partial L_e} + Q_2 \frac{\partial P_2}{\partial L_e} \quad , \ < \text{ only if } L_e^* = 0$$

Therefore

$$(5\text{-}4\text{-}6) \quad \eta_y \leq \theta_1 \eta_{P_1} + \theta_2 \eta_{P_2} \quad , \ < \text{ only if } L_e^* = 0$$

where $\eta_y \equiv \frac{\partial y}{\partial L_e} \cdot \frac{L_e}{y}$ is the elasticity of after-tax income with respect to enforcement, $\theta_i \equiv P_i Q_i / y$ is the share of after-tax income spent on commodity i, and $\eta_{P_i}$ is the elasticity of the regulated equilibrium price of commodity i with respect to enforcement.

Clearly the extreme solution $L_e = L$ is never optimal since production of both commodities, and hence after-tax income and utility, would then be zero. This possibility is then excluded from consideration in the following analysis.

These results are now summarized.

<u>PROPOSITION 5-4-1</u>: <u>If an externality is regulated by an output quota which is enforced by fines, then, assuming an interior solution, the optimal level of enforcement according to NPIT equates the rate of commodity substitution to the rate of product transformation through regulatory enforcement. Equivalently, the optimal level of enforcement equates its expenditure-share weighted impact on prices to its impact on after-tax income. If the rate of substitution in</u>

consumption of commodity 2 for commodity 1 is everywhere exceeded by the rate of product transformation through enforcement, or alternatively if the expenditure-share weighted impact on prices of enforcement everywhere exceeds its impact on after-tax income, then optimal enforcement is zero. It is never optimal to devote the entire labour force to enforcement activity.

As shown by (5-4-2) and (5-4-4) it is possible that it is not desirable within the NPIT framework to attempt to enforce the regulations. A sufficient condition for enforcement to be welfare improving and thus initiated by a NPIT regulator is that

$$(5\text{-}4\text{-}7) \qquad \frac{u_2}{u_1} > - \left. \frac{\partial Q_2/\partial L_e}{\partial Q_1/\partial L_e} \right|_{L_e = 0}$$

which is the case as illustrated in Figure 5-3-2. This condition is not necessary as it is theoretically possible, depending on the properties of the regulated equilibrium locus, for efficiency gains at higher levels of enforcement to outweigh initial reductions in utility.

NPIT regulation is now examined in the Cobb-Douglas case assuming the enforcement process is also Ricardian so that $f(L_e) \equiv L_e$. Note that

$$(5\text{-}4\text{-}8) \qquad L_e \leq \bar{L}_e < L$$

where $\bar{L}_e$ is the level of labour devoted to enforcement which generates a probability of detection of unity. It is clear that if $L_e = L$ all offences are detected since none occur. It seems reasonable to assume that all offences would be detected if the enforcement effort was large relative to the offence-generating labourforce and therefore $\bar{L}_e < L$. Further increases in enforcement labour beyond the level $\bar{L}_e$ represent a pure waste of resources and, as such, will not be undertaken by a NPIT regulator.

In the Cobb-Douglas case the indirect utility function is

$$(5\text{-}4\text{-}9) \qquad V(L_e) = v(P_1, P_2, y) = \frac{Ky}{P_1^b P_2^{1-b}}$$

where $K = b^b(1-b)^{1-b}$ is a positive constant depending on b. Substituting for regulated prices and income from (5-3-4), (5-3-22) and (5-3-23) and using the assumption $f(L_e) \equiv L_e$ gives

$$(5\text{-}4\text{-}10) \qquad V(L_e) = \frac{K[L-L_e+a_3L_eQ_1]}{[a_1+a_3L_e]^b[a_2(J)]^{1-b}}$$

where

$$(5\text{-}4\text{-}11) \qquad J(a_3, L_e) = \frac{a_1b(L-L_e)}{a_1+(1-b)a_3L_e}$$

is the regulated equilibrium employment level in industry 1 in the Cobb-Douglas case given (5-2-1) and (5-3-31).

The level of social welfare at the unregulated competitive equilibrium sets a lower bound on the level of social welfare for the economy under NPIT regulation for no regulation will be enforced, under NPIT criteria, if it fails to at least maintain welfare at its unregulated level.

A sufficient condition for enforcement to be welfare improving in the Cobb-Douglas is that

(5-4-12) $\quad V'(0) > 0$

Substituting for $Q_1$ in the numerator of (5-4-10) gives

(5-4-13) $$V(L_e) = \frac{K(a_1+a_3L_e)(L-L_e)}{[a_1+a_3L_e]^b[a_2(J)]^{1-b}[a_1+(1-b)a_3L_e]}$$

As shown in Appendix 4-1, differentiating (5-4-13) with respect to enforcement and much manipulation reveals that

(5-4-14) $$V'(L_e) \gtrless 0 \text{ if and only if } \xi(L_e) \gtrless B(L_e);\ 0 \leq L_e < \bar{L}_e$$

where

(5-4-15) $$\xi(L_e) = \frac{a_2'(J)L_1}{a_2(J)}$$

is the elasticity of the input-output coefficient in industry 2, and

(5-4-16) $$B(L_e) = \frac{1}{1-b} - \frac{a_3[L-L_e](a_1+(1-b)a_3L_e)}{(a_1+a_3L_e)(a_1+(1-b)a_3L)}$$

Examination of the properties of $B(L_e)$ in Appendix 5-2 shows that

(5-4-17) $$B(0) = \frac{a_1}{(1-b)(a_1+(1-b)a_3L)}\ ,\ B'(L_e) > 0$$

From Proposition 5-3-2 $L_1$ is decreasing in $L_e$. Therefore $\xi(L_e)$ is non-increasing in $L_e$ if and only if $\xi(L_e)$ is non-decreasing in $L_1$. This observation together with (5-4-17) shows that $V'(L_e) < 0$ for all $L_e \geq 0$ if

(5-4-18) $$\xi(0) = \frac{a_2'(bL)bL}{a_2(bL)} < \frac{a_1}{(1-b)(a_1+(1-b)a_3L)} = B(0)$$

where $\xi(0)$ is the elasticity of the input-output coefficient in industry 2 evaluated at the unregulated competitive equilibrium and $bL$ is the unregulated competitive equilibrium employment level in industry 1 given by $J(a_3,0)$ from (5-4-11).

Expression (5-4-18) then places a lower bound on the elasticity term for enforcement to occur under NPIT regulation.

LEMMA 5-4-1

If $\xi(L_e)$ is non-increasing in $L_e$, then $V'(0) > 0$ is a necessary and sufficient condition for enforcement to occur under a NPIT regulator.

Proof

Enforcement occurs under NPIT regulation if and only if it raises aggregate welfare above the unregulated competitive equilibrium level. Expression (5-4-12) gives a sufficient condition for welfare to be raised by enforcement. A necessary condition is that there exists $L_e \in (0,\bar{L}_e)$ such that $V'(L_e) > 0$.

From (5-4-14) $V'(L_e) > 0$ requires that $\xi(L_e) > B(L_e)$. From (5-4-17) it is known that $B(L_e)$ is strictly increasing in $L_e$. If $\xi(L_e)$ is non-increasing in $L_e$, which requires that (A5-2-13) holds, then, from the monotonicity of both functions, $\xi(L_e)$ cannot exceed $B(L_e)$ at any level of enforcement unless it does so at $L_e = 0$.

Therefore $\xi(0) > B(0)$ is a necessary and, by (5-4-12), sufficient condition for NPIT enforcement to occur in the Cobb-Douglas case given the assumption that $\xi(L_e)$ is non-increasing in $L_e$. □

It is instructive to examine how the condition given in (5-4-18) is affected by changes in the parameters b and L.

By hypothesis $\xi(0)$ is non-decreasing in L whereas $B(0)$ is strictly decreasing in L. The derivation of $\partial B(0)/\partial L$ and the necessary condition to support the assumption on $\xi(0)$ are given in Appendix 5-3.

Following from this, there exists a unique $\hat{L}$ such that

(5-4-19) $\quad \xi(0) \lesseqgtr B(0)$ if and only if $L \gtreqless \hat{L}$

Finally, as shown in Appendix 5-4, as a function of b, B(0) is positive, increasing and $B(0) \to \infty$ as $b \to 1$ while $\xi(0)$ is finite and assumed to be non-decreasing. From (A5-4-2) and (A5-4-7) in Appendix 5-4

(5-4-20) $\quad \xi(0)\Big|_{b=0} = 0 < B(0)\Big|_{b=0}$

while from (A5-4-3) and (A5-4-8)

(5-4-21) $\quad \lim_{b\to 1} B(0) = \infty > \xi(0)\Big|_{b=1}$

Let $\tilde{b}$ be the smallest value of b and $\hat{b}$ the largest volume of b such that $\xi(0) = B(0)$. Then

(5-4-22) $\quad \xi(0) < B(0) \quad \forall\, b\colon \quad 0 \leq b < \tilde{b}$ and $\hat{b} < b \leq 1$

If therefore there exists any $b \in (0,1)$ such that $\xi(0) > B(0)$ then

(5-4-23) $\quad b \in [\tilde{b},\hat{b}] \ ; \ \tilde{b} \leq \hat{b}$

and enforcement is welfare improving for such b under the NPIT hypothesis given Cobb-Douglas preferences. In summary:

<u>PROPOSITION 5-4-2</u>: <u>Suppose that the elasticity of the input-output coefficient in industry 2 is non-decreasing as activity increases in industry 1, that the probability of detection is linear in enforcement, and that preferences are Cobb-Douglas. Enforcement is only carried out if this elasticity is sufficiently large</u>.

Moreover, there is a unique $\hat{L}$ such that resources are devoted to enforcement if and only if the labour force exceeds $\hat{L}$. There are unique $\tilde{b}$, $\hat{b}$, $\tilde{b} \leq \hat{b}$ such that no resources are devoted to enforcement if $b < \tilde{b}$ or $b > \hat{b}$.

All three parts of Proposition 5-4-2 are intuitively appealing. Equations (5-4-3) and (5-4-13) are functions relating the level of utility to the amount of labour devoted to enforcement which, given the values of the parameters b and L, provide the regulatory agency with complete information as to how its enforcement policies affect social welfare for any binding quota. Underlying these equations is the model structure that any regulation involves a tradeoff between efficiency and resource capacity. Under the NPIT hypothesis, regulation will not occur unless this tradeoff is beneficial for society as a whole. That is, the increase in utility which results from the shift towards efficient production must outweigh the reduction in utility from the lower productive capacity of the economy which results from the diversion of labour to enforcement activities. Thus in Figure 5-3-2 E' lies on a higher social indifference curve than the unregulated competitive equilibrium at E: even though the production possibility set is diminished, NPIT regulation is welfare improving.

In accordance with this argument, the first result of Proposition 5-4-2 shows that regulation raises social welfare above its competitive equilibrium level if and only if the externality generated by industry 1 is sufficiently severe to warrant incurring the resource costs of enforcement. This is significant not only when considering the responsiveness of productivity in industry 2 to changes in the output of commodity 1 but also with respect to the relative sizes of the industries at the unregulated competitive equilibrium. As shown by (5-4-22), if industry 1 is relatively small then the efficiency gains from restricting its output will be insufficient to offset the costs of so doing.

Secondly, (5-4-19) shows that if the economy is small it is too costly to devote resources to enforcement of a regulatory control on the externality generating industry. The strength of this result stems from the form of the detection function which here is solely dependent on the amount of labour devoted to enforcement. In the analysis in Chapter Two it was shown that different forms of penalty function exhibit different characteristics. The caveat was stressed,however,that all enforcement activity requires the use of scarce resources and thus the qualitative implications of the present result remain.

Finally, as shown by expression (5-4-21), if industry 1 is sufficiently important to consumers it ought not to be regulated within the NPIT framework. This occurs as $b \to 1$ and the relative size of industry 1 at the unregulated competitive equilibrium increases. Conversely, the externality generating industry should be regulated only if the industry which it adversely affects is important to consumers and the negative externality is sufficiently severe.

## 5-5 REGULATION ACCORDING TO CT

According to the Naive Public Interest Theory the level of the quota, as given in (5-3-16), is of no concern to regulators other than the consideration of an appropriate upper bound. This is because the marginal expected penalty used is independent of the extent to which production exceeds the quota and so is independent of the quota itself. As a consequence, the marginal expected fine is equivalent to a non-avoidable unit tax on the output of the externality-generating industry. It is this which restricts the industry's output. Since the tax rate is independent of the size of the quota the regulator is happy as long as the quota is binding. If the regulators have explicit distributional goals these will be achieved without regard to the quota and its enforcement in the NPIT case.

As shown in Section 5-3, with reference to Figure 5-3-1, the level and allocation of quotas has significant distributional impacts. All individuals will then be concerned about the distribution of the quotas. The essential hypothesis of CT is that members of the regulated industry will gain control of the regulatory agency and operate it in their own interests. The distribution, size, and enforcement, of the quota are then mutually determined. This behaviour is motivated by the following objective.

It is assumed that the CT regulator acts to maximize the utility of those agents employed in the regulated industry. The determination of equilibrium prices within the regulatory environment was discussed in Section 5-3. Before the problem faced by the regulator can be formulated, the income of the interest group must be defined.

The after-tax income of employees in industry 1 is

$$(5\text{-}5\text{-}1)\qquad y_1(a_3,L_e,R_1,R) = L_1 + a_3f(L_e)R_1 + a_3f(L_e)[Q_1-R]\frac{L_1}{L} - L_e\frac{L_1}{L}$$

where $L_1$ is the income from employment in the industry, $R_1$ is the allocation of quota to firms within the industry, and $a_3f(L_e)R_1$ is the imputed rental value of this quota. An additional source of income is the share of these workers in the fines imposed by the regulatory agency. If fines are redistributed in lumpsum fashion, which is consistent with the assumption made concerning funding of the enforcement effort, the term $a_3f(L_e)[Q-R]\frac{L_1}{L}$ is added to the interest group's income. Finally, the term $L_e\frac{L_1}{L}$ is subtracted to allow for the lumpsum tax which pays for the regulatory agency's activities.[2]

Recalling that the regulated equilibrium output of industry 1 is a function of $a_3$ and $L_e$, so too, from (5-2-1), is $L_1$.

The indirect utility function can be used to evaluate the utility of the employees in industry 1. Let

$$(5\text{-}5\text{-}2)\qquad V_1(a_3,L_e,R_1,R)=v(P_1(a_3,L_e),P_2(a_3,L_e),y_1(a_3,L_e,R_1,R))$$

The objective of the CT regulator depends on the level of enforcement as well as the size and distribution of the quota. Here, as in the analysis of NPIT - optimal regulation, the unit fine $a_3$ is taken as given. Note that neither $R_1$ nor R affect the prices. Also

$$(5\text{-}5\text{-}3)\qquad 0 \le R_1 \le R \le Q_1(a_3,L_e)$$

since the quota allocated to the industry cannot exceed the total quota, while the quota is valuable only if it is not greater than the regulated equilibrium output of commodity 1. Differentiating (5-5-2) with respect to $R_1$ using (5-5-1) and (5-5-3) gives

$$(5\text{-}5\text{-}4) \qquad \frac{\partial V_1}{\partial R_1} = \frac{\partial v}{\partial y_1} . a_3 f(L_e) > 0$$

To maximize the utility of the regulated group then $R_1 = R$. That is, the captured regulator will award itself all of the quota. Using this gives

$$(5\text{-}5\text{-}5) \qquad \frac{\partial V_1}{\partial R} = \frac{\partial v}{\partial y_1} . a_3 f(L_e)[\frac{L-L_1}{L}] > 0$$

Therefore $R = Q_1(a_3, L_e)$. To maximize utility of those employed in industry 1 the CT regulator will set the quota equal to the regulated equilibrium output. If this is not done, then some fines will be collected, and only a part of the proceeds will be redistributed to the captors of the regulatory agency. To avoid this dilution of the quota's value it is set so that it is just observed.

Even though the quota is not exceeded, enforcement of the regulation is still carried out. This is because the rents which accrue to the quota holders depend on enforcement.

In view of (5-5-4) and (5-5-5), and using (5-2-1), (5-5-1) simplifies to

$$(5\text{-}5\text{-}6) \qquad \tilde{y}_1(a_3, L_e) = [a_1 + a_3 f(L_e) - \frac{a_1 L_e}{L}] Q_1(a_3, L_e)$$

Define

$$(5\text{-}5\text{-}7) \qquad \tilde{V}_1(L_e) = v(P_1(a_3, L_e), P_2(a_3, L_e), \tilde{y}_1(a_3, L_e))$$

This is maximized by choosing $L_e$. The first-order condition for an interior maximum yields,

$$\eta_{\tilde{y}_1} = \theta_1 \eta_{P_1} + \theta_2 \eta_{P_2} \tag{5-5-8}$$

where $\eta_{\tilde{y}_1} \equiv \frac{\partial \tilde{y}_1}{\partial L_e} \cdot \frac{L_e}{\tilde{y}_1}$ is the elasticity of after-tax income for employees of industry 1 with respect to enforcement. The remaining variables in (5-5-8) are as in (5-4-6) which gives the optimality condition in the NPIT case.

These results are summarized in the following Proposition.

PROPOSITION 5-5-1: A captured regulatory agency will set quotas on the externality-generating industry equal to its output. No offences against the regulations will occur, so that no fines will be collected. All of the quota will be assigned to the industry. Labour will be assigned to enforcement so that its expenditure-weighted impact on prices equals its impact on the income of the employees of the externality-generating industry.

The behaviour of the CT regulator is now examined assuming that preferences are represented by the Cobb-Douglas direct utility function given in (5-2-19).

The corresponding indirect utility function for labour employed in the regulated industry is

$$\tilde{V}_1(L_e) = v\,(P_1, P_2, \tilde{y}_1) = \frac{K\tilde{y}_1}{P_1^b P_2^{1-b}} \tag{5-5-9}$$

where $K = b^b(1-b)^b$ is a positive constant.

Substituting from (5-3-4), (5-3-22) and (5-5-6) and using the assumption $f(L_e) \equiv L_e$ gives

$$(5\text{-}5\text{-}10) \quad \tilde{V}_1(L_e) = \frac{K[a_1(L-L_e)+a_3LL_e]b(L-L_e)}{(a_1+a_eL_e)^b[a_2(J)]^{1-b}[a_1+(1-b)a_3L_e]L} \quad ; \quad L_e \leq \bar{L}_e$$

The unregulated competitive equilibrium in the Cobb-Douglas case is characterized by (5-2-21). The associated utility level for the employees in industry 1 sets a lower bound on their utility under CT regulation for no regulation will be enforced under CT criteria if it fails to at least maintain the original welfare level within the industry.

A sufficient condition for CT regulation to occur is that

$$(5\text{-}5\text{-}11) \qquad \tilde{V}_1'(0) > 0$$

As shown in Appendix 5-5, differentiating (5-5-10) and much manipulation reveals that

$$(5\text{-}5\text{-}12) \quad \tilde{V}_1'(L_e) \gtrless 0 \text{ if and only if } \xi(L_e) \gtrless B_1(L_e) \quad ; \quad 0 \leq L_e < \bar{L}_e$$

where

$$(5\text{-}5\text{-}13) \quad B_1(L_e) = \left[\frac{1}{1-b}\right] \cdot \left[1 + \frac{(L-L_e)[a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e)+(1-b)a_1+a_3L_e)]}{(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e-(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]}\right] \left[\frac{[a_1(L-L_e)+a_3LL_e-(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]}{[a_1(L-L_e)+a_3LL_e][a_1+(1-b)a_3L]}\right]$$

From Appendix 5-5 and using (A5-2-6)

$$(5\text{-}5\text{-}14)\quad B_1(0) = \frac{2a_1}{(1-b)(a_1+(1-b)a_3L)}\ ,\ B_1(L_e) > 0\ ;\ 0 \leq L_e < L$$

and

$$(5\text{-}5\text{-}15)\quad \lim_{L_e \to L} B_1(L_e) = \lim_{L_e \to L} B(L_e) = \frac{1}{1-b}$$

Using (5-5-15) in (5-5-11) gives a sufficient condition for CT regulation that

$$(5\text{-}5\text{-}16)\quad \xi(0) = \frac{a_2(bL)bL}{a_2(bL)} > \frac{2a_1}{(1-b)(a_1+(1-b)a_3L)} = B_1(0)$$

Comparison of the right-hand side of (5-4-17) and (5-5-16) shows that

$$(5\text{-}5\text{-}17)\qquad B_1(0) = 2B(0)$$

The condition for regulation to be undertaken is therefore more stringent in the CT case than for the NPIT regulator. Unfortunately, examination of (A5-6-4) in Appendix 5-6 does not yield any unambiguous results about the slope of $B_1(L_e)$. Whereas the converse of (5-4-18) is necessary and sufficient for enforcement to occur in the NPIT case, assuming that $\xi(L_e)$ is non-increasing in $L_e$, it is theoretically possible for enforcement to occur uncer a CT regulator even though the condition given by (5-5-16) is not satisfied.

<u>PROPOSITION 5-5-2</u>: <u>Suppose that the elasticity of the input-output coefficient in industy 2 is non-decreasing as activity in industry 1 increases, that the probability of detection is linear in enforcement, and that preferences are Cobb-Douglas. Enforcement is carried out only if this elasticity is</u>

sufficiently large. There is a unique $\hat{\hat{L}}$ such that resources are devoted to enforcement only if the labour force exceeds $\hat{\hat{L}}$. There are unique $\tilde{\tilde{b}}$, $\hat{\hat{b}}$, $\tilde{\tilde{b}} \leq \hat{\hat{b}}$ such that no resources are devoted to enforcement if $b < \tilde{\tilde{b}}$ or $b > \hat{\hat{b}}$.

Proof.

Enforcement occurs within CT regulation if and only if it improves the welfare of those employed in the regulated industry compared with the level they enjoy at the unregulated competitive equilibrium.

From (5-5-12) then, a necessary condition for CT enforcement is that there exists some $L_e > 0$ such that

(5-5-18) $\xi(L_e) > B_1(L_e)$

From (A5-6-1) and (A5-6-2)

(5-5-19) $B_1(L_e) > B(L_e) \quad \forall\, L_e, \quad 0 \leq L_e \leq \bar{L}_e$

Therefore, using (5-5-18) and (5-5-19), a necessary condition for CT enforcement is that there exists some $L_e > 0$ such that

(5-5-20) $\xi(L_e) > B(L_e)$

That is, the possibility of NPIT enforcement is necessary for CT enforcement to occur.

By Lemma (5-4-1) the condition expressed in (5-5-20) is satisfied if and only if $\xi(0) > B(0)$ assuming that $\xi(L_e)$ is non-increasing in $L_e$.

Therefore, a necessary condition for CT enforcement to occur assuming that $\xi(L_e)$ is non-increasing in $L_e$ is that

(5-5-21) $\xi(0) > B(0)$

Expression (5-5-21) is the converse of (5-4-18) and hence a lower bound is placed on the elasticity term to admit the possibility of enforcement occurring in CT regulation.

From (5-4-19), (5-5-21) is satisfied if and only if $L > \hat{L}$ where $\hat{L}$ is the size of the labour force such that $\xi(0) = B(0)$. This ensures that NPIT enforcement will occur when $L > \hat{L}$. Using (5-5-19), however, (5-5-18) is not satisfied in the neighbourhood of $\hat{L}$ and hence for enforcement to occur in the CT case, it is necessary that the labourforce exceed $\hat{\hat{L}}$ where $\hat{\hat{L}} > \hat{L}$.

From (5-4-22), (5-5-21) is not satisfied for any $b \leq \tilde{b}$ or $b \geq \hat{b}$ where $\tilde{b}$ is the smallest value of b and $\hat{b}$ the largest value of b such that $\xi(0) = B(0)$. Using (5-5-19), however, (5-5-18) is not satisfied in the neighbourhood of $\tilde{b}$ and $\hat{b}$ and therefore no resources will be devoted to CT enforcement if $b < \tilde{\tilde{b}}$ or $b > \hat{\hat{b}}$ where $\tilde{\tilde{b}} \geq \tilde{b}$ and $\hat{\hat{b}} \leq \hat{b}$, $\tilde{\tilde{b}} \leq \hat{\hat{b}}$.

□

These results are again intuitive. With a small labourforce, enforcement costs, of which the interest group bears a proportionate share, are prohibitive relative to the benefits, while as the relative size of the industry increases, the potential for garnishing further resources from the rest of the economy diminishes. If, however, the industry is relatively small, there is little scope for rental income from quotas and small likelihood of a favourable effect on the inter-sectoral terms of trade.

## 5-6 A COMPARISON OF NPIT AND CT REGULATION

The previous sections have established results concerning the behaviour of a regulatory agency under two different assumptions about its objective. In each case the agency operates in the same economic environment and has available to it the same policy instruments. Differences in the settings of these instruments, which are the level and distribution of the output quota and the labour devoted to regulatory enforcements, arise from the differences in the agency's objective.

The most visible sign that the regulatory agency is captured is that production will never exceed the quota, or, equivalently, that no fines are ever imposed on the externality-generating industry. The agency will be concerned about the size of the quota, and will seek to distribute it to members of the regulated industry. In contrast, if the public interest is being served by the regulations then the agency will not be concerned about the size of the quota, other than with regard to the question of an upper bound. Fines will be imposed on the externality-generating industry for it will in general exceed its quota. Since it is likely that the regulators' distributional goals include equity, in all probability the quota will not be exclusively assigned to the regulated industry.

Although (5-4-6) and (5-5-8), which define the optimal enforcement under NPIT and CT, are similar they do not yield the same solution. Denote by $\hat{L}_e$ the NPIT-optimal enforcement, and by $\tilde{L}_e$ the CT-optimal enforcement, of the regulations.

Equations (5-3-6) and (5-5-6) give the after tax regulated incomes of the respective interest groups under NPIT and CT regulation. Using $Q_1(a_3,L_e)$ and differentiating gives

$$(5\text{-}6\text{-}1) \qquad \frac{\partial y}{\partial L_e} = a_3 f'(L_e)Q_1 + a_3 f(L_e)\frac{\partial Q_1}{\partial L_e} - 1$$

and

$$(5\text{-}6\text{-}2) \qquad \frac{\partial \tilde{y}_1}{\partial L_e} = [a_3 f'(L_e) - \frac{a_1}{L}]Q_1 + [a_1 + a_3 f(L_e) - \frac{a_1 L_e}{L}]\frac{\partial Q_1}{\partial L_e}$$

Comparing (5-6-1) and (5-6-2)

$$(5\text{-}6\text{-}3) \qquad \frac{\partial \tilde{y}_1}{\partial L_e} = \frac{\partial y}{\partial L_e} + a_1[1 - \frac{L_e}{L}]\frac{\partial Q_1}{\partial L_e} - \frac{a_1}{L} Q_1(a_3, L_e) + 1$$

From the second order condition for a maximum it follows that

$$(5\text{-}6\text{-}4) \qquad \tilde{L}_e \lesseqgtr \hat{L}_e \text{ if and only if } \left.\frac{\partial \tilde{y}_1}{\partial L_e}\right|_{\hat{L}_e} \lesseqgtr \left.\frac{\partial y}{\partial L_e}\right|_{\hat{L}_e}$$

Equation (5-4-2) implies that $\partial Q_1/\partial L_e < 0$ at the NPIT optimum.

Rearranging and simplifying (5-6-3) using (5-2-1) and applying the result in (5-6-4) reveals that

$$(5\text{-}6\text{-}5) \qquad \tilde{L}_e \lesseqgtr \hat{L}_e \text{ if and only if}$$

$$\eta(L_e) \lesseqgtr - \frac{(L-L_1)L_e}{(L-L_e)L_1} \quad ; \quad L_e = \hat{L}_e \; ; \quad \tilde{L}_e, \hat{L}_e \geq 0$$

where $\eta(L_e) = \dfrac{\partial Q_1}{\partial L_e} \dfrac{L_e}{Q_1(a_3,L_e)}$ is the elasticity of regulated equilibrium output in industry 1 with respect to changes in the level of enforcement.

This result is intuitively appealing. As the discussion in Section 5-3 showed, with reference to Figure 5-3-1, enforcement acts to increase the market price of the regulated commodity which in the CT case is captured by the regulated industry in the form of imputed quota rentals. If at the NPIT-optimal level of enforcement $\hat{L}_e$ the effect of changes in enforcement on prices and incomes within the regulated environment is such that equilibrium output in industry 1 is relatively inelastic, with respect to $L_e$, then it is optimal for the CT regulator to expand the enforcement effort. If, however, $Q_1(a_3, L_e)$ is relatively elastic, the opposite holds true. The CT regulator is essentially acting as a monopolist seeking, among other things, to maximize the rental value of the quota.

The use of resources to enforce the quota within the CT framework has as its purpose the assurance of quota rents to members of the regulated industry. It is possible therefore, from (5-6-5), that optimal enforcement under CT regulation could exceed that under NPIT to such an extent that social welfare is reduced below its unregulated competitive equilibrium level even though the welfare of those employed in the regulated industry rises.

From (5-6-5), however, if the NPIT-optimal enforcement level $\hat{L}_e$ is zero and it is assumed that $\eta(L_e)$ is continuous then $\eta(0) = 0$ and the optimal level of enforcement under CT is also zero. If therefore $\eta(L_e)$ is continuous in $L_e$, the possibility of NPIT enforcement is necessary for enforcement to occur under CT regulation.

These results are summarized by the following Proposition.

<u>PROPOSITION 5-6-1:</u> <u>A captured regulatory agency will choose an output quota just equal to the industry output, and assign this quota</u>

<u>to members of the industry. If the agency is maximizing social welfare it will not care about the size of the quota, and its allocation will be more general. No fines will be imposed by a captured agency, although fines would be expected if the agency maximizes social welfare. Depending on the elasticity of regulated equilibrium output in industry 1 with respect to enforcement, evaluated at the NPIT optimal level of enforcement, the CT regulator may devote more or less rsources to enforcing the regulation than is socially optimal. Social welfare under CT regulation cannot exceed that under NPIT regulation. If the elasticity of regulated equilibrium output is continuous then enforcement will not occur under CT regulation unless enforcement would also occur under a NPIT regulator.</u>

Comparing the behaviour of the NPIT and CT regulators in the Cobb-Douglas framework gives the following results.

<u>PROPOSITION 5-6-2:</u> <u>Suppose that the elasticity of the intput-output coefficient in industry 2 is non-decreasing as activity in industry 1 increases, that the probability of detection is linear in</u>

<u>enforcement, and that preferrences are Cobb-Douglas. The NPIT optimal level of enforcement exceeds the CT optimal level. CT regulation does not reduce social welfare below its unregulated competitive equilibrium level. The labourforce required for enforcement to occur in the CT case is larger than that under NPIT regulation. The criteria on the size of the industry in the unregulated competitive equilibrium that admit the possibility of enforcement are more stringent in the CT case than under NPIT regulation</u>.

<u>Proof</u>

As shown in the proof of Proposition 5-4-2, if $\xi(L_e)$ is non-decreasing in $L_1$, there is a unique NPIT optimal level of enforcement

(5-6-6) $$\hat{L}_e \quad ; \quad 0 < L_e < \bar{L}_e$$

provided that the converse of (5-4-17) holds. At this enforcement level, (5-4-13), (5-4-14) and (5-4-15) show that

(5-6-7) $$\xi(L_e) = B(\hat{L}_e)$$

Given the assumptions on $\xi(L_e)$ and $B(L_e)$ with respect to changes in enforcement levels

(5-6-8) $$V'(L_e) < 0 \qquad \forall\ L_e > \hat{L}_e$$

Expressions (A5-6-1) and (A5-6-2) in Appendix 5-6 show that

$$(5\text{-}6\text{-}9) \qquad B_1(L_e) > B(L_e) \quad ; \quad 0 \leq L_e < \bar{L}_e$$

Using this result in (A5-5-6) reveals therefore that

$$(5\text{-}6\text{-}10) \qquad \tilde{V}_1'(L_e) < 0 \qquad \forall\, L_e \geq \hat{L}_e$$

Expression (5-6-10) implies that if enforcement is profitable within the CT framework, then

$$(5\text{-}6\text{-}11) \qquad 0 \leq \tilde{L}_e < \hat{L}_e$$

where $\tilde{L}_e$ is the CT optimal level of enforcement.

The remaining results follow from the proof of Proposition 5-5-2.

□

## 5-7 CONCLUSION

This chapter has shown, within the particular context of a Ricardian economy with a negative externality generating industry faced with an output quota enforced by means of a constant marginal expected monetary penalty, that the results derived in Chapters Two and Three in the partial equilibrium context extend to a general equilibrium approach.

The observable differences in behaviour which arise from each hypothesis provide, in principle, a way of discerning the objectives of the regulator from its actions. Most important among these are the results concerning the size and distribution of the quota which affect expected penalty payments and rental income.

In addition it has been shown that the CT regulator chooses the quota size, distribution and enforcement level, so as to maximize its rental value allowing for the resource cost of enforcement and its effect on relative prices.

This provides an extension within the free-entry general equilibrium context of the result derived in Chapter Three showing that, in a partial equilibrium framework, the CT regulator forms a cartel-like situation in the regulated industry.

Finally, the NPIT and CT optimal enforcement levels differ according to the elasticity of regulated equilibrium output in the regulated industry with respect to changes in the level of enforcement. In particular it is possible to conceive of a situation where CT regulation reduces social welfare below that associated with the unregulated competitive equilibrium.

If, however, the elasticity of the regulated equilibrium output of the externality-generating industry is continuous, then the existence of enforcement under CT regulation implies the potential for NPIT enforcement which will, in general, raise social welfare. This result together with that which shows the unambiguous rise in social welfare that occurs with CT regulation in the Cobb-Douglas case suggests that a doctrinaire approach to deregulation is misleading and potentially damaging to the economy.

While deregulation may improve social welfare when the regulatory agency is captured, the possibility exists, and seems more likely, that a return to the unregulated competitive equilibrium will reduce aggregate welfare. Any attempt to deregulate an industry therefore should proceed only after a careful evaluation of the operation of the regulation concerned.

More important than blanket emphasis on deregulation is monitoring the performance of the regulatory agency.

This makes the ability to infer the objectives of a regulator from its actions especially significant. With resource costs of enforcing a regulation the first-best optimum, shown by G in Figure 5-2-2,is unattainable. Any NPIT regulation increases social welfare. The NPIT optimal level of regulatory activity, which may or may not coincide with the unregulated competitive equilibrium, then produces the second-best optimum for the economy.

In general, CT regulation inevitably results in a lower level of social welfare than at the NPIT optimum. If CT regulation can be detected from observations of the actions of the regulatory agency, the operation of the regulation can be transformed to conform with NPIT objectives. This process will likely increase social welfare and cannot reduce it provided that the costs of so doing are small relative to the potential welfare gains.

In any event, it would appear that the informational requirements and costs of the performance monitoring process are similar to those of a well-considered deregulation programme while the potential benefits are greater. This suggests a policy of monitoring and improving the performance of regulatory agencies, rather than deregulation, if it is desired to increase aggregate welfare in a regulated environment.

NOTES

* This chapter is based on a joint paper with R. Manning entitled "A simple general equilibrium comparison of theories of regulation and enforcement" (unpublished). The paper has been presented at the New Zealand Economists' Association Conference, Dunedin, August 1985, the Reserve Bank of New Zealand Seminar Series, Wellington, November 1985, and the Australasian Econometrics Society Conference, Melbourne, August 1986. Murray Kemp and Ross Wilson are thanked for helpful comments.

1. An interesting interpretation of this tax is that each individual is required to devote part of their labour to policing the regulation. This gives each agent an incentive to check on the compliance of others with the aims of the regulatory agency. Law enforcers are thus kept honest which, as shown by Becker and Stigler [1975] is a problem for any regulatory system.
2. The formulation of regulated after-tax income in industry 1 given in (5-5-1) reflects the free-entry assumption of atomistic individual firms within the industry. Each firm then faces the expected penalty cost of breaking the law but perceives only an arbitrarily small amount of this being returned in the proportional reallocation of fine revenue. The unit trading price of the quota is then the marginal expected penalty $a_3 f(L_e)$.

   If, however, an industry association was acquiring quotas on the open market, its demand price from equation (4-5-6) would be $a_3 f(L_e)[\frac{L-L_1}{L}]$ which represents the net cost of illegal production. Depending on the allocation of quotas outside the industry , a bilateral monopoly may emerge with a supply price for quotas at $a_3 f(L_e)$. The trading price of the quota would then depend on the relative strengths of the parties in the bargaining process.

In any event, neither assumption affects the central result that a maximizing CT regulator will award the industry all of the quota at a level such that no fines occur at the regulated equilibrium.

Appendix 5-1

Differentiating (5-4-12) with respect to enforcement labour gives

$$(A5\text{-}1\text{-}1)\quad V'L_e = \Big[K[a_1+a_3L_e]^b[a_2(J)]^{1-b}[a_1+(1-b)a_3L_e]\big[a_3(L-L_e)$$
$$-(a_1+a_3L_e)\big] - K(a_1+a_3L_e)(L-L_e)\Big[ba_3[a_1+a_3L_e]^{b-1}$$
$$[a_2(J)]^{1-b}[a_1+(1-b)a_3L_e] + (1-b)a_3[a_1+a_3L_e]^b[a_2(J)]^{1-b}$$
$$+ [a_1+a_3L_e]^b[a_1+(1-b)a_3L_e]\Big[(1-b)[a_2(J)]^{-b}a_2'(J)\cdot$$
$$\frac{[-a_1b[a_1+(1-b)a_3L]]}{[a_1+(1-b)a_3L_e]^2}\Big]\Big]\Big] \Big/$$
$$\Big[[a_1+a_3L_e]^b[a_2(J)]^{1-b}[a_1+(1-b)a_3L_e]\Big]^2$$

As the denominator is positive, the sign of the derivative is given by the sign of the numerator. Recalling (5-4-8), the range of enforcement levels considered is restricted to $L_e$ $[0,\bar{L}_e]$. Simplifying the numerator of (A5-1-1) then gives

$$(A5\text{-}1\text{-}2)\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$
$$a_2(J)[a_1+(1-b)a_3L_e][a_3(L-L_e)-(a_1+a_3L_e)]$$
$$-(L-L_e)\Big[ba_3[a_1+(1-b)a_3L_e]a_2(J) + (1-b)a_3[a_1+a_3L_e]a_2(J)$$
$$-a_1b(1-b)(a_1+a_3L_e)\left[\frac{[a_1+(1-b)a_3L]}{a_1+(1-b)a_3L_e}\right]a_2'(J)\Big] \gtreqless 0;\quad 0 \leq L_e < \bar{L}_e$$

Rearranging (A5-1-2)

$$(A5\text{-}1\text{-}3)\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$
$$a_2(J)\Big[[a_1+(1-b)a_3L_e][a_3(L-L_e)-(a_1+a_3L_e)]$$
$$-b(L-L_e)a_3[a_1+(1-b)a_3L_e] - (1-b)(L-L_e)a_3[a_1+a_3L_e]\Big]$$
$$+a_2'(J)\Big[a_1b(1-b)(L-L_e)[a_1+a_3L_e]\frac{[a_1+(1-b)a_3L]}{a_1+(1-b)a_3L_e}\Big] \gtreqless 0$$

Rearranging the term in $a_2(J)$ gives

$$\text{(A5-1-4)}\quad a_2(J)\Big[[a_1+(1-b)a_3L_e][a_3(L-L_e)-ba_3(L-L_e)] - (a_1+a_3L_e)[(1-b)a_3(L-L_e)+a_1+(1-b)a_3L_e]\Big]$$

and simplifying

$$\text{(A5-1-5)}\quad a_2(J)\Big[(1-b)a_3(L-L_e)[a_1+(1-b)a_3L_e] - (a_1+a_3L_e)[a_1+(1-b)a_3L]\Big]$$

Substituting (A5-1-5) into (A5-1-3) gives

$$\text{(A5-1-6)}\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$

$$a_2(J)\Big[(1-b)a_3(L-L_e)[a_1+(1-b)a_3L_e] - (a_1+a_3L_e)[a_1+(1-b)a_3L]\Big] + a_2'(J)\left[a_1b(1-b)(L-L_e)[a_1+a_3L_e]\frac{[a_1+(1-b)a_3L]}{a_1+(1-b)a_3L_e}\right] \gtreqless 0$$

Further simplification reveals

$$\text{(A5-1-7)}\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$

$$a_2(J)\left[\frac{a_3(L-L_e)[a_1+(1-b)a_3L_e]}{[a_1+(1-b)a_3L](a_1+a_3L_e)} - \frac{1}{1-b}\right] + a_2'(J)\left[\frac{a_1b(L-L_e)}{a_1+(1-b)a_3L_e}\right] \gtreqless 0$$

Rearranging (A5-1-7) gives

$$\text{(A5-1-8)}\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$

$$\frac{a_2'(J)}{a_2(J)} \gtrless \left[\frac{1}{1-b} - \frac{a_3(L-L_e)[a_1+(1-b)a_3L_e]}{(a_1+a_3L_e)[a_1+(1-b)a_3L]}\right]\left[\frac{a_1+(1-b)a_3L_e}{a_1b(L-L_e)}\right]$$

Multiplying both sides of (A5-1-8) by $L_1$ and using (5-3-32) gives

$$\text{(A5-1-9)}\quad V'(L_e) \gtreqless 0 \text{ if and only if}$$

$$\frac{a_2'(J)L_1}{a_2(J)} \gtrless \left[\frac{1}{1-b} - \frac{a_3(L-L_e)[a_1+(1-b)a_3L_e]}{(a_1+a_3L_e)[a_1+(1-b)a_3L]}\right];\ 0 \leq L_e < \bar{L}_e$$

The left-hand side of (A5-1-9) is the elasticity of the input-output coefficient in industry 2 denoted by $\xi(L_e)$ in (5-4-15) while the right-hand side is the function $B(L_e)$ as shown in the text by (5-4-16).

Appendix 5-2

Expanding the function $B(L_e)$ from (5-4-16) gives

(A5-2-1) $B(L_e) =$

$$\frac{(a_1+a_3L_e)[a_1+(1-b)a_3L]-(1-b)a_3(L-L_e)[a_1+(1-b)a_3L_e]}{(1-b)(a_1+a_3L_e)[a_1+(1-b)a_3L]}$$

As the denominator is positive for all $L_e \geq 0$ the sign of (A5-2-1) depends on that of the numerator. Expanding further

(A5-2-2) $B(L_e) \gtreqless 0$ if and only if

$$a_1^2+(1-b)a_1a_3L + a_1a_3L_e + (1-b)a_3^2LL_e$$

$$+ (1-b)a_3L_e[a_1+(1-b)a_3L_e] - (1-b)a_1a_3L$$

$$- (1-b)(1-b)a_3^2LL_e \gtreqless 0$$

Simplifying terms and substituting in (A5-2-1) gives

(A5-2-3) $B(L_e) =$

$$\frac{a_1(a_1+a_3L_e)+b(1-b)a_3^2LL_e+(1-b)a_3L_e[a_1+(1-b)a_3L_e]}{(1-b)(a_1+a_3L_e)[a_1+(1-b)a_3L]}$$

$$> 0 \;\forall\; L_e \;;\; 0 \leq L_e < \bar{L}_e$$

Evaluating (A5-2-3) at $L_e = 0$ gives

(A5-2-4) $$B(0) = \frac{a_1^2}{(1-b)a_1[a_1+(1-b)a_3L]}$$

which simplifying yields (5-4-17) in the text.

Recalling the derivation of (5-4-16) in Appendix 5-1

(A5-2-5) $\lim_{L_e \to L} B(L_e) =$

$$\frac{a_1(a_1+a_3L) + b(1-b)a_3^2L^2 + (1-b)a_3L[a_1+(1-b)a_3L]}{(1-b)(a_1+a_3L)[a_1+(1-b)a_3L]}$$

Simplifying and collecting terms gives

(A5-2-6) $$\lim_{L_e \to L} B(L_e) = \frac{(a_1+a_3L)[a_1+(1-b)a_3L]}{(1-b)(a_1+a_3L)[a_1+(1-b)a_3L]} = \frac{1}{1-b}$$

This result may be seen by direct inspection of (5-4-16).

Differentiating (5-4-16) with respect to $L_e$ gives

$$(A5\text{-}2\text{-}7) \quad B'(L_e) = -\Big[(a_1+a_3L_e)[a_1+(1-b)a_3L]\cdot \Big[(1-b)a_3^2(L-L_e) - a_3[a_1+(1-b)a_3L_e]\Big] - a_3(L-L_e)[a_1+(1-b)a_3L_e]a_3[a_1+(1-b)a_3L]\Big] \Big/ [(a_1+a_3L_e)[a_1+(1-b)a_3L]]^2$$

The sign of (A5-2-7) depends on that of the numerator. Rearranging

$$(A5\text{-}2\text{-}8) \quad B'(L_e) \gtreqless 0 \text{ if and only if}$$

$$a_3[a_1+(1-b)a_3L]\Big[\overset{(i)}{(1-b)a_3(a_1+a_3L_e)[L-L_e]} - \overset{(ii)}{[a_1+(1-b)a_3L_e]} - \overset{(iii)}{a_3(L-L_e)[a_1+(1-b)a_3L_e]}\Big] \gtreqless 0$$

Term (ii) in (A5-2-8) is clearly negative for all $L_e > 0$. Taking terms (i) and (iii) a sufficient condition is

$$(A5\text{-}2\text{-}9) \quad B'L_e) \gtreqless 0 \text{ if}$$

$$a_3(L-L_e)\Big[(1-b)(a_1+a_3L_e) - [a_1+(1-b)a_3L_e]\Big] \lesseqgtr 0$$

Simplifying the bracketed term gives

$$(A5\text{-}2\text{-}10) \quad (1-b)a_1 + (1-b)a_3L_e - a_1 - (1-b)a_3L_e = -ba_1 < 0$$

Substituting into (A5-2-8) and (A5-2-9) shows that $B'(L_e) > 0$ for all $L_e$, $0 \leq L_e \leq \bar{L}_e$.

Differentiating the elasticity term (5-4-15) with respect to $L_1$

$$(A5\text{-}2\text{-}11) \quad \frac{\partial}{\partial L_1}\left[\frac{a_2'(L_1)L_1}{a_2(L_1)}\right]$$

$$= \Big[a_2(L_1)[a_2'(L_1)+a_2''(L_1)L_1] - a_2'(L_1)L_1a_2'(L_1)\Big] \Big/ [a_2(L_1)]^2$$

The sign of (A5-2-11) depends on that of the numerator. Therefore

(A5-2-12) $\frac{\partial}{\partial L_1}[\xi(L_e)] \gtrless 0$ if and only if

$$a_2''(L_1)L_1a_2(L_1) + a_2'(L_1)a_2(L_1) \gtrless [a_2'(L_1)]^2L_1$$

Simplifying and rearranging reveals that the necessary condition for $\xi(L_e)$ to be non-decreasing in $L_1$ is

(A5-2-13) $$a_2''(L_1) \geq \frac{a_2'(L_1)}{a_2(L_1)L_1}[a_2'(L_1)L_1-a_2(L_1)] \ , \ L_1 > 0.$$

Appendix 5-3

Differentiating the elasticity term in (5-4-18) with respect to L

$$(A5\text{-}3\text{-}1) \quad \frac{\partial}{\partial L}\left[\frac{a_2'(bL)bL}{a_2(bL)}\right] = \Big[a_2(bL)[ba_2''(bL)bL+ba_2'(bL)]$$

$$- a_2'(bL)bLa_2'(bL)b\Big]/[a_2(bL)]^2$$

The sign of (A5-3-1) depends on that of the numerator. Therefore

$$(A5\text{-}3\text{-}2) \quad \frac{\partial}{\partial L}[\xi(0)] \gtreqless 0 \text{ if and only if}$$

$$a_2''(bL)a_2(bL)b^2L + a_2(bL)ba_2'(bL) \gtreqless [a_2'(bL)]^2b^2L$$

Simplifying and rearranging reveals that

$$(A5\text{-}3\text{-}3) \quad \frac{\partial}{\partial L}[\xi(0)] \gtreqless 0 \text{ if and only if}$$

$$a_2''(bL) \gtreqless \frac{[a_2'(bL)]^2}{a_2(bL)} - \frac{a_2'(bL)}{bL}$$

From (A5-3-3) a necessary condition for $\xi(0)$ to be non decreasing in L is that

$$(A5\text{-}3\text{-}4) \quad a_2''(bL) \geq \frac{a_2'(bL)}{a_2(bL)bL}[a_2'(bL)bL - a_2(bL)]$$

From (5-2-9), using (5-2-1) and (5-2-2) and substituting for the competitive equilibrium employment levels in both industries, the necessary condition for the curvature of the production possibility frontier as portrayed in Figure 5-2-1 is

$$(A5\text{-}3\text{-}5) \quad a_2''(bL) < \frac{a_2'(bL)}{(1-b)L}\left[1 + a_2'(bL)\frac{(1-b)L}{a_2(bL)}\right]$$

In order to satisfy both conditions, it is necessary that

$$(A5\text{-}3\text{-}6) \qquad \frac{a_2'(bL)}{a_2(bL)bL}\,[a_2'(bL)bL - a_2(bL)]$$

$$< \frac{a_2'(bL)}{a_2(bL)(1-b)L}\,[a_2(bL) + a_2'(bL)(1-b)L]$$

Rearranging, (A5-3-6) gives

$$(A5\text{-}3\text{-}7) \qquad \frac{1-b}{b}\left[a_2'(bL)bL - a_2(bL) - a_2(bL)\frac{1}{1-b} - a_2(bL)(1-b)L\frac{1}{1-b}\right] < 0$$

Simplifying (A5-3-7) gives

$$(A5\text{-}3\text{-}8) \qquad -\frac{a_2(bL)}{b} < 0$$

and the condition given in (A5-3-6) holds.

To satisfy both the hypothesis on $\xi(0)$ and the assumed curvature property it is necessary that

$$(A5\text{-}3\text{-}9) \qquad \frac{a_2'(bL)}{a_2(bL)bL}\,[a_2'(bL)bL - a_2(bL)] \leq a_2''(bL)$$

$$< \frac{a_2'(bL)}{(1-b)L}\;1 + a_2'(bL)\frac{(1-b)L}{a_2(bL)}$$

Differentiating B(0) with respect to L gives

$$(A5\text{-}3\text{-}10) \qquad \frac{\partial}{\partial L}\,[B(0)] = \frac{-(1-b)^2 a_3}{[(1-b)[a_1+(1-b)a_3L]]^2} < 0$$

and B(0) is obviously decreasing in L as assumed in the text.

Appendix 5-4

Differentiating the right hand side of (5-4-18) with respect to b gives

$$(A5\text{-}4\text{-}1) \quad \frac{\partial}{\partial b}[B(0)] = \frac{-a_1[-[a_1+(1-b)a_3L]-a_3L(1-b)]}{[(1-b)[a_1+(1-b)a_3L]]^2} > 0$$

which is obviously positive as stated in the text.

Evaluating the function at its end-points

$$(A5\text{-}4\text{-}2) \quad B(0)\Big|_{b=0} = \frac{a_1}{a_1+a_3L} > 0$$

and

$$(A5\text{-}4\text{-}3) \quad \lim_{b\to 1} B(0) = \frac{a_1}{0[a_1]} = \infty$$

Differentiating the left hand side of (5-4-18) with respect to b gives

$$(A5\text{-}4\text{-}4) \quad \frac{\partial}{\partial b}[\xi(0)]$$

$$= \frac{a_2(bL)[a_2'(bL)L + a_2''(bL)bL^2] - a_2'(bL)bL^2 a_2'(bL)}{[a_2(bL)]^2}$$

The sign of (A5-4-4) depends on that of the numerator. Therefore

$$(A5\text{-}4\text{-}5) \quad \frac{\partial}{\partial b}[\xi(0)] \gtrless 0 \text{ if and only if}$$

$$a_2(bL)[a_2'(bL)L + a_2''(bL)bL^2] \gtrless [a_2'(bL)]^2bL^2$$

Simplifying and rearranging reveals that a necessary condition for $\xi(0)$ to be non-decreasing in b is

$$(A5\text{-}4\text{-}6) \quad a_2''(bL) \geq \frac{a_2'(bL)}{a_2(bL)bL}[a_2'(bL)bL - a_2(bL)]$$

Expression (A5-4-6) is exactly the same as (A5-3-4) and therefore the same conclusions apply.

Evaluating the function at its end-points

$$(A5\text{-}4\text{-}7) \quad \xi(0)\Big|_{b=0} = \frac{a_2'(0)0}{a_2(0)} = 0$$

where $a_2(0)$ is the value of the input-output coefficient in industry 2 in the absence of externality.

$$(A5\text{-}4\text{-}8) \quad \xi(0)\Big|_{b=1} = \frac{a_2'(L)L}{a_2(L)} > 0$$

Appendix 5-5

Differentiating (5-5-10) with respect to $L_e$ gives

(A5-5-1) $\tilde{V}_1'(L_e)$

$$= \Bigg[ KL\Big[(a_1+a_3L_e)^b (a_2(J))^{1-b}[a_1+(1-b)a_3L_e]\Big]\Big[-b[a_1(L-L_e) + a_3LL_e] + b(L-L_e)[a_3L-a_1]\Big] - KL\Big[b(L-L_e)[a_1(L-L_e) + a_3LL_e]\Big]\Big[b(a_1+a_3L_e)^{b-1}a_3(a_2(J))^{1-b}(a_1+(1-b)a_3L_e) + (1-b)a_3(a_1+a_3L_e)^b(a_2(J))^{1-b} + (a_1+a_3L_e)^b(a_1+(1-b)a_3L_e)(1-b)(a_2(J))^{-b}a_2'(J)\cdot \left[\frac{-ba_1(a_1+(1-b)a_3L)}{(a_1+(1-b)a_3L_e)^2}\right]\Big]\Bigg] \Big/ \Big[(a_1+a_3L_e)^b(a_2(J))^{1-b}[a_1 + (1-b)a_3L_e]L\Big]^2$$

The sign of the derivative is given by the sign of the numerator. The range of enforcement levels as in Appendix 5-1 is restricted to $L_e \in [0,\bar{L}_e]$. Simplifying the numerator of (A5-5-1) gives

(A5-5-2) $\tilde{V}_1'(L_e) \gtrless 0$ if and only if

$$(a_2(J))(a_1+a_3L_e)[a_1+(1-b)a_3L_e]\Big[b(L-L_e)[a_3L-a_1] - b[a_1(L-L_e)+a_3LL_e]\Big] - \Big[b(L-L_e)[a_1(L-L_e) + a_3LL_e]\Big]\cdot \Big[ba_3(a_2(J))(a_1+(1-b)a_3L_e) + (1-b)a_3(a_2(J))(a_1+a_3L_e) - a_1(a_1+a_3L_e)(a_2'(J))b(1-b)\frac{(a_1+(1-b)a_3L)}{a_1+(1-b)a_3L_e}\Big] \gtrless 0$$

Collecting terms in $a_2(J)$ and rearranging gives

(A5-5-3) $\tilde{V}_1'(L_e) \gtrless 0$ if and only if

$$(a_2(J))(a_1+a_3L_e)(a_1+(1-b)a_3L_e)\Big[\Big[b(L-L_e)(a_3L-a_1)$$

$$- b[a_1(L-L_e)+a_3LL_e]\overset{(i)}{]} - b(L-L_e)[a_1(L-L_e) + a_3LL_e] \cdot$$

$$\overset{(ii)}{\frac{[ba_3(a_1+(1-b)a_3L_e) + (1-b)a_3(a_1+a_3L_e)]}{(a_1+a_3L_e)(a_1+(1-b)a_3L_e)}} \Bigg]$$

$$+ a_2'(J)b(L-L_e)[a_1(L-L_e) + a_3LL_e] \cdot$$

$$\left[\overset{(iii)}{b(1-b)a_1(a_1+a_3L_e)}\frac{(a_1+(1-b)a_3L)}{(a_1+(1-b)a_3L_e)}\right] \gtrless 0$$

Simplifying term (i) the expression becomes

$$\text{(A5-5-4)} \quad a_2(J)\left[-\frac{1}{1-b} - \left[\frac{b(L-L_e)[a_1(L-L_e)+a_3LL_e]}{[b[a_1(L-L_e)+a_3LL_e]-b(L-L_e)(a_3L-a_1)]}\right] \cdot \right.$$

$$\left.\left[\frac{ba_3(a_1+(1-b)a_3L_e) + (1-b)a_3(a_1+a_3L_e)}{(1-b)[a_1+a_3L_e][a_1+(1-b)a_3L_e]}\right]\right]$$

$$+ a_2'(J)\frac{b(L-L_e)[a_1(L-L_e)+a_3LL_e]ba_1(a_1+(1-b)a_3L)}{[b[a_1(L-L_e)+a_3LL_e]-b(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]^2}$$

Rearranging (A5-5-4) and simplifying gives

$$\text{(A5-5-5)} \quad \tilde{V}_1'(L_e) \gtrless 0 \text{ if and only if } \frac{a_2'(J)}{a_2(J)} \gtrless \frac{1}{1-b} \cdot$$

$$\left[1 + \frac{b(L-L_e)[a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e)+(1-b)a_3(a_1+a_3L_e)]}{(a_1+a_3L_e)[b[a_1(L-L_e)+a_3LL_e]-b(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]}\right] \cdot$$

$$\left[\frac{[b[a_1(L-L_e)+a_3LL_e] - b(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]^2}{b(L-L_e)[a_1(L-L_e)+a_3LL_e]ba_1[a_1+(1-b)a_3L]}\right];$$

$$0 \leq L_e < \bar{L}_e$$

Multiplying both sides of (A5-5-5) by $L_1$ and using (5-3-32) gives

$$\text{(A5-5-6)} \quad \tilde{V}_1'(L_e) \gtrless 0 \text{ if and only if } \frac{a_2'(J)L_1}{a_2(J)} \gtrless \frac{1}{1-b} \cdot$$

$$\left[1 + \frac{b(L-L_e)[a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e)+(1-b)a_3(a_1+a_3L_e)]}{(a_1+a_3L_e)[b[a_1(L-L_e)+a_3LL_e]-b(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]}\right]$$

$$\left[\frac{[b[a_1(L-L_e)+a_3LL_e] - b(L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]}{b[a_1(L-L_e)+a_3LL_e][a_1+(1-b)a_3L]}\right];$$

$$0 \leq L < \bar{L}$$

The left-hand side of this expression is the elasticity of the input-output coefficient in industry 2 as denoted by $\xi(L_e)$ in (5-4-15) and (5-5-12) while the right-hand side is the function $B_1(L_e)$ shown in the text by (5-5-13).

The right hand side of (A5-5-6) can alternatively be expressed as follows.

$$\text{(A5-5-7)} \quad B_1(L_e) = \Big[\overset{\text{(i)}}{(a_1+a_3L_e)}\,[\overset{\text{(ii)}}{[a_1(L-L_e)+a_3LL_e]-(L-L_e)(a_3L-a_1)}]\cdot$$

$$[a_1+(1-b)a_3L_e] + \overset{\text{(iii)}}{(L-L_e)[a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e)}$$

$$+ (1-b)a_3(a_1+a_3L_e)]\Big] \,/\, [(1-b)(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e]$$

$$(a_1+(1-b)a_3L)]$$

Using (A5-5-7) and simplifying

$$\text{(A5-5-8)} \quad B_1(0) = \frac{2a_1}{(1-b)(a_1+(1-b)a_3L)}$$

while from the right-hand side of (A5-5-6)

$$\text{(A5-5-9)} \quad \lim_{L_e \to L} B_1(L_e) = \frac{1}{1-b}[1+0]\left[\frac{ba_3L^2(a_1+(1-b)a_3L)}{ba_3L^2(a_1+(1-b)a_3L)}\right] = \frac{1}{1-b}$$

From (A5-5-7), as term (i) is positive for all $L_e \geq 0$, upon rearranging terms (ii) and (iii) it is sufficient for $B_1(L_e) > 0$ that

$$\text{(A5-5-10)} \quad [a_1(L-L_e) + a_3LL_e][ba_3(a_1+(1-b)a_3L_e)$$

$$+ (1-b)a_3(a_1+a_3L_e)] + a_1(a_1+a_3L_e)(a_1+(1-b)a_3L_e)$$

$$- a_3L(a_1+a_3L_e)(a_1+(1-b)a_3L_e) > 0$$

Expanding (A5-5-10)

(A5-5-11) $ba_1^2a_3L + b(1-b)a_1a_3^2LL_e - ba_1^2a_3L_e - b(1-b)a_1a_3^2L_e^2$

$+ (1-b)a_1^2a_3L + (1-b)a_1a_3^2LL_e - (1-b)a_1^2a_3L_e - (1-b)a_1a_3^2L_e^2$

$+ ba_1a_3^2LL_e + b(1-b)a_3^3LL_e^2 + (1-b)a_1a_3^2LL_e$

$+ (1-b)a_3^3LL_e^2 + a_1^3 + (1-b)a_1^2a_3L_e + a_1^2a_3L_e + (1-b)a_1a_3^2L_e^2$

$- a_1^2a_3L - a_1a_3^2LL_e - (1-b)a_1a_3^2LL_e - (1-b)a_3^3LL_e^2$

Collecting terms and simplifying gives

(A5-5-12) $b(1-b)a_3^3LL_e^2 + b(1-b)a_1a_3^2(L-L_e)L_e + (1-b)a_1^2a_3L_e$

$+ a_1^3 > 0 \qquad \forall\, L_e \geq 0$

Using (A5-5-10), (A5-5-12) then reveals that $B_1(L_e)$ is strictly positive for all $L_e$, $0 \leq L_e < \bar{L}_e$.

Appendix 5-6

Using (A5-2-1) and (A5-5-7) and simplifying

(A5-6-1) $B_1(L_e) \gtreqless B(L_e)$ if and only if

$$(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e - (L-L_e)(a_3L-a_1)](a_1+(1-b)a_3L_e)$$
$$+ (L-L_e)[a_1(L-L_e)+a_3LL_e][a_3(a_1+(1-b)a_3L_e)+(1-b)a_3(a_1+a_3L_e)]$$
$$- (a_1+a_3L_e)(a_1+(1-b)a_3L)[a_1(L-L_e) + a_3LL_e] \gtreqless 0$$

Rearranging gives

(A5-6-2) $$(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e][a_1+(1-b)a_3L_e-(a_1+(1-b)a_3L)]$$
$$+ (L-L_e)[a_1(L-L_e)+a_3LL_e][(1-b)a_3(a_1+a_3L_e)]$$
$$+ (L-L_e)(a_1+(1-b)a_3L_e)[a_3[a_1(L-L_e)+a_3LL_e]$$
$$- (a_3L-a_1)(a_1+a_3L_e)]$$

Simplifying (A5-6-2) reveals

(A5-6-3) $(L-L_e)(a_1+(1-b)a_3L_e)[a_1^2] > 0 \;\forall\; L_e \;\; 0 \leq L_e < L$

which, using (A4-6-1), shows that $B_1(L_e) > B(L_e)$ provided that the entire labourforce is not devoted to enforcement activities; a possibility that has previously been excluded.

Differentiating (A5-5-7) with respect to $L_e$ gives

(A5-6-4) $$B_1'(L_e) = (1-b)(a_1+(1-b)a_3L)\Big[(a_1+a_3L_e)[a_1(L-L_e)$$
$$+ a_3LL_e]\Big[a_3[a_1(L-L_e)+a_3LL_e - (L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]$$
$$+ (1-b)a_3(a_1+a_3L_e)[a_1(L-L_e) + a_3LL_e - (L-L_e)(a_3L-a_1)]$$
$$+ (a_1+a_3L_e)(a_1+(1-b)a_3L_e)[2(a_3L-a_1)]$$
$$- [a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e) + (1-b)a_3(a_1+a_3L_e)]$$
$$+ (L-L_e)(a_3L-a_1)[ba_3(a_1+(1-b)a_3L_e) + (1-b)a_3(a_1+a_3L_e)]$$
$$+ (L-L_e)[a_1(L-L_e) + a_3LL_e][b(1-b)a_3^2 + (1-b)a_3^2]\Big]$$
$$-\Big[(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e - (L-L_e)(a_3L-a_1)][a_1+(1-b)a_3L_e]$$

$$+ (L-L_e)[a_1(L-L_e)+a_3LL_e][ba_3(a_1+(1-b)a_3L_e)+(1-b)a_3(a_1 + a_3L_e]\Big][a_3[a_1(L-L_e)+a_3LL_e]+(a_1+a_3L_e)(a_3L-a_1)]\Big] / [(1-b)(a_1+(1-b)a_3L)(a_1+a_3L_e)[a_1(L-L_e)+a_3LL_e]]^2$$

Unfortunately much tedious manipulation of the expression fails to yield any unambiguous results about its sign.

# CHAPTER SIX

# CONCLUSION

Regulation is an area of economic activity that has had, and continues to receive, extensive coverage in the literature. Many theories have been developed in an attempt to explain the existence and practice of regulation. However, as was evidenced in Chapter One where several such theories were reviewed, no single theory explains the observed contrasts in regulatory behaviour and no clear criteria exist by which to identify a particular regulation as corresponding to one theory or another. This, it was argued, results from the lack of a consistent theoretical basis linking the various approaches together. The construction of such a basis was one of the aims of the present thesis.

Until recently, theoretical models of regulation proceeded as if compliance with regulatory constraints was perfect and voluntary. The recognition that this is not the case in practice leads to the need to model the enforcement process. Each of the models developed here therefore incorporated a formal theoretical treatment of the enforcement process with enforcement being taken as neither costless nor complete.

Accordingly, the analytical section in Chapter Two began by developing a partial equilibrium model of a competitive negative-externality generating industry in which output was regulated by means of either an output quota or unit rate sales tax each of which was enforced by means of an expected monetary penalty dependent on the size of output and/or constraint violation and the level of resources devoted to enforcement activity. These examples were chosen in light of the dispute in the literature concerning the relative merits of "price" and "quantitative" regulatory instruments.

A regulated equilibrium was defined within the model as the point of equation between market demand and market supply which emerges as the result of individual optimization within the regulatory environment. Given the assumption of the model that the industry comprised a finite number of identical competitive firms, the regulated equilibrium was shown to occur at that output level where the marginal expected penalty equated the difference between the demand price and marginal cost of production. This result, which is one of the main results of the enforcement literature, was found to be dependent on the form of the expected penalty function. In particular, some functional forms, such as the commonly legislated maximum penalty, were shown to severely limit the deterrence capabilities of the regulator.

The enforcement-induced marginal expected penalty acts as an unavoidable tax on output and it is this "tax" which supports the existence of the constraint and which generates the regulated equilibrium. The marginal expected penalty may be viewed as the "price" of engaging in illegal activity. This is equally true whether regulation is by sales tax or output quota. Indeed, it was shown that it is possible to design tax/expected penalty and quota/

expected penalty combinations with identical aggregate incomes. When illegal behaviour and enforcement are allowed for then, there is a sense in which any regulatory control works fundamentally through the price mechanism. It was also shown that increases in enforcement, assuming that the expected penalty remained binding on behaviour, by raising the marginal expected penalty at any level of illegal output and associated extent of constraint violation, reduce the regulated equilibrium output level of the industry. This is the well-known deterrence result. Ranging across all feasible enforcement levels, the associated output levels define the locus of regulated equilibria for the industry which, in a partial equilibrium model, corresponds to the market demand curve at various quantities not exceeding the unregulated competitive equilibrium output level.

Given that any output level not exceeding the unregulated competitive equilibrium is technically feasible with some non-negative expected penalty, and that the aggregate outcome is independent of the regulatory instrument used, it is evident that the regulatory policy employed is dependent on something other than a simple output objective. If it can be determined that the characteristics of a particular regulatory policy correspond to a particular aspect of the regulatory process, then, in principle it is possible to infer the existence of this aspect of the regulatory process from the observation of the policy characteristics. The assertion of this thesis was that regulatory behaviour could be explained in terms of the objectives of the regulator. The second aim of the thesis was then to establish a theoretical structure by which it was possible to consistently explain variations in regulatory behaviour, particularly in enforcement practice, and to, in turn, infer regulatory objectives from observed behaviour.

Two competing hypotheses concerning regulatory objectives were incorporated into the otherwise identical model of regulation developed in Chapter Two. These were the Naive Public Interest Theory (NPIT) under which the regulator is assumed to seek the maximization of aggregate social welfare, and the Capture Theory (CT) which, in the pure sense as used here, holds that the regulator acts so as to maximize the profit of the regulated industry. In practice, CT regulation was shown to be somewhat akin to the cartelization of the industry with the function of the marginal expected penalty being to solve the internal chiselling problem inherent in the operation of a cartel.

The analysis derived and compared the characteristics of optimal regulatory policy under each hypothesis and also examined the effects of changes in various relevant parameters on these policies. It was found that, in general, optimal policy differed between the two regulators and that these policy differences were dependent on the form of the expected penalty function. The source of many of the differences was the contrasting distributional aspects of the regulators' objectives. A NPIT regulator, concerned solely with the aggregate outcome, is indifferent to the distributional effects of the various policies whereas the behaviour of a CT regulator, concerned only with a subset of economic agents, is largely determined by them.

In the case of an expected penalty function that was strictly convex in both output and the extent of constraint violation and contained no fixed rate component, it was shown that a NPIT regulator would not employ a non-zero output quota but would regulate the industry by means of a unit rate sales tax, the rate of which was no less than the marginal expected penalty with respect to output at the NPIT-optimal regulated equilibrium output level. This result was motivated by the necessity to minimize the enforcement cost of achieving a given output level in the regulated environment. The CT regulator, however, preferred to control the industry by means of a non-zero output quota allocated entirely to members of the regulated industry.

This contrast in the choice of regulatory instrument provides, within the context of the model, a clearly observable criterion by which to infer the objectives of a regulator by its actions. Namely it can be concluded that, with an expected penalty function that is strictly convex in both output and the extent of constraint violation and contains no fixed rate component, regulation by means of a unit rate sales tax is inspired by NPIT objectives whereas a non-zero output quota reveals the existence of a captured regulator.

A similar result also holds under the assumption of a flat rate per unit expected penalty. With this penalty structure the marginal expected penalty is independent of the extent of the violation. Any change in the size of available quota, provided that it remains binding, has no effect on the regulated equilibrium output level. In this case it was shown that the NPIT regulator was indifferent between a sales tax, the rate of which was not less than the per unit expected penalty, and any binding output quota. The CT regulator again preferred an output quota to a sales tax. As in the case of the strictly convex expected penalty function, this output quota was allocated entirely to members of the regulated industry. In this case, however, the CT-optimal policy had the additional characteristic that the quota was set exactly at the CT-optimal regulated equilibrium output level. With no quota violations, and hence no incurred fines, this policy has the feature that all expected penalty payments accrue to members of the regulated industry in the form of implicit quota rentals.

Other differences in policy between NPIT and CT regulators, while no less real, are not readily identifiable from observed behaviour. Differences were shown to exist in optimal levels of enforcement activity and in the responses of the regulators to changes in various parameters such as the level of aggregate consumer income and the state of enforcement technology. In each of these cases it was not possible to infer the objectives of the regulator from the observation of its behaviour without possessing enough a priori information to solve the optimization problem of each type of regulator. Of all the parameter changes examined it was possible to readily distinguish behaviour only in the case of those with purely distributional effects. Thus, for instance, a change in the level of enforcement activity in response to a change in the proportion of the resource cost of enforcement funded by the industry was shown to reveal the existence of a captured regulator.

The majority of the partial equilibrium analysis dealt with the regulation of a competitive industry but the case of a monopolistic industry was touched upon briefly. Unregulated monopolistic control of the industry was assumed to result in an output level which, although less than the corresponding output level in the competitive case, remained socially excessive. In these circumstances it was shown that NPIT-optimal regulation, if it occurred, was on a scale much reduced from that in the competitive case. Any enforcement-induced reduction in output from the pure-monopoly profit-maximizing equilibrium level lowers industry profit and thus would not be undertaken voluntarily by a CT regulator. This raised the possibility of strategic regulation whereby a CT regulator might institute some small degree of control over the industry in an effort to preclude more hostile NPIT regulation. It was concluded that anything other than token restrictions was an indication that the monopoly was not being controlled by an unfettered CT regulator.

The theoretical framework of regulation developed in the context of a negative-externality generating industry was then applied to the case of an open-access fishery. Using Copes' [1970, 1972] long run steady-state industry supply curve which combines factor cost and population stock/yield relationships, together with a flat rate per unit expected penalty, the analysis showed that the model of regulation developed in the thesis is not specific to any particular market structure or regulatory context.

The Copes' supply curve is backward-bending over a certain output-price range reflecting the fact that any steady-state harvest, other than the maximum sustainable yield output level, is associated with two stock sizes and, through the technical production function, with two levels of factor input also. A distinction was drawn between fisheries in which the unregulated open-access equilibrium was assumed to occur on the upward-sloping and backward-sloping portions of the industry supply curve. These were referred to as "low-cost" and "high-cost" fisheries respectively. The one peculiarity of the "high-cost" fishery case was the perverse "deterrence" result whereby, through the interaction of population dynamics and production cost considerations, an increase in the enforcement of a binding output quota resulted in an increased regulated equilibrium output level and hence a greater degree of illegal behaviour. With the exception of this, and the fact that the nature of the common property externality which exists in an open-access fishery leads to the optimal degree of control over a "low-cost" fishery for a CT regulator exceeding that of a NPIT regulator, the results of this analysis echo those in the case of the regulation of a negative-externality generating industry.

As stated in the review of the literature, the majority of theoretical analyses of regulation are of a partial equilibrium nature. Concentration on a particular industry or sector of an economy in isolation precludes consideration of the inter-sectoral impacts of regulation and the implications, in turn, for the operation of regulatory policy of its effects elsewhere in the economy. In order to examine such considerations and implications, the model of the regulation of a negative-externality generating industry was incorporated within a simple Ricardian two sector general equilibrium model of an economy with labour as the only factor of production, again using an output quota enforced by means of a flat rate per unit expected penalty function. Aside from the more general results, a Cobb-Douglas utility function was taken to represent consumer preferences as an illustrative device to examine the characteristics of regulatory policy under more particular assumptions.

After the unregulated competitive equilibrium within the general equilibrium framework was defined, the operation of regulation within the model was explained and the locus of regulated equilibria derived. This locus, which coincides with the unregulated competitive equilibrium when no enforcement is carried out, and approaches the origin as the entire resources of the economy are devoted to enforcement activities, reinforces the fact that regulation is a costly business and graphically illustrates the tradeoff that exists between the reduction in the productive capacity of the economy as a result of enforcement activity and the efficiency gains from the enforcement-induced contraction in the output of the externality-generating industry. The slope of this locus depends on the interaction between production technology, demand conditions, and enforcement activity. In particular, if the regulated commodity is inferior, it was shown that it was possible for an increase in enforcement activity to lead to an increase in regulated equilibrium output and hence an increase in the extent of constraint violation also. The deterrence result therefore does not necessarily hold in a general equilibrium framework even with a form of expected penalty function for which it does in a partial equilibrium context. Non-inferiority was shown to be a sufficient condition to generate the deterrence result. This was illustrated for the case of Cobb-Douglas preferences.

The locus of regulated equilibria contains all feasible competitive equilibria that can be generated within the regulatory environment. The point on this locus at which the economy is positioned is determined by the objectives of the regulator. Given the possibility of inaction, enforcement will occur only if it increases the value of the regulator's objective function. In general equilibrium, the conditions for viability are intuitively appealing. Enforcement is more likely to be welfare improving the larger is the labour force, (the larger is the labour force the smaller is the proportionate reduction in the economy's productive resources from a unit of labour devoted to enforcement activities), the greater is the degree of externality, and the more moderately sized is the regulated industry. (A large industry implies that the commodity it produces is too highly valued by consumers to restrict output despite the adverse effect it has on other industries while a small industry does not generate enough damage to make regulation worthwhile.)

Assuming that enforcement is viable, the NPIT-optimal regulated equilibrium was found to occur at the point where the slope of the regulated equilibrium locus, which shows the rate at which one commodity can be transformed into another through the regulatory process, was tangential to a community indifference curve. This point will not, in general, coincide with the optimal regulated equilibrium of the CT regulator seeking to maximize the welfare of members of the regulated industry only. It was shown that the relative amounts of enforcement that were optimal for the two regulators depended on the elasticity of the output of the regulated commodity in response to enforcement. The more inelastic is the regulated commodity to increases in enforcement, the more it is likely that optimal enforcement under a CT regulator exceeds that for a NPIT regulator. It is possible, therefore, that CT-optimal enforcement could exceed that under NPIT regulation to such an extent that aggregate welfare is reduced below its unregulated competitive equilibrium level even though the welfare of those employed in the regulated industry rises. Additionally, however, it was found that CT regulation could not profitably take place unless NPIT regulation was also viable. In the case of CT regulation, therefore, the possibility of a welfare non-decreasing transformation to NPIT-optimal regulation always exists.

As was illustrated in the case of Cobb-Douglas preferences, there are circumstances in which CT regulation does not reduce aggregate welfare below its unregulated competitive equilibrium level. Combining this with the immediately preceding result suggests that a policy of deregulation per se could therefore be detrimental to aggregate welfare. Of greater benefit would be the ability to determine the objectives of a regulator so that regulatory policy could then be amended as appropriate. To this end the observable characteristics of optimal regulatory policies found in the partial equilibrium framework under the assumption of a flat rate per unit expected penalty were shown to hold in the general equilibrium context also. Specifically it was again demonstrated that a captured regulator would set the output quota at the regulated equilibrium output level and allocate it entirely to members of the regulated industry allowing all penalty payments to be captured by the industry in the form of implicit quota rentals.

This thesis then has constructed a simple theoretical model of the regulation of a competitive industry explicitly incorporating the necessity for enforcement. The concept of a regulated equilibrium was defined and, by ranging across all feasible levels of enforcement activity, the locus of all regulated equilibria determined. By providing an explanation of the operation of the regulatory process, the model provides a consistent theoretical basis by which to explain the theory and practice of regulation and on which particular instances of regulation can be compared. It was demonstrated that the model is not specific to any particular regulatory situation and is applicable in both partial and general equilibrium contexts.

Into this otherwise identical model of the regulatory process, two hypotheses concerning the objectives of the regulator were embedded and the behavioural characteristics of optimal regulatory policy derived for each case. Comparison of the results in both partial and general equilibrium contexts revealed clear contrasts between the policies of the two regulators, some of which were readily identifiable. The fact that the optimal regulated equilibrium of a particular regulator is associated with certain observable characteristics implies that, in principle, the objectives of a regulator can be inferred from the observation of regulatory practice. Given the relationship demonstrated in the general equilibrium context between aggregate welfare levels at the unregulated competitive equilibrium and at the regulated equilibria associated with each type of regulator, the existence of clear and observable criteria by which to determine a regulator's objectives would enable the monitoring and modifying as necessary of regulatory behaviour so as to optimize regulatory performance over time in terms of society's objectives.

There are several areas in which the analysis presented here could be usefully extended. These include the application of the model to other areas of "market failure" such as external economies of scale and the incorporation of further regulatory objective hypotheses such as Niskanen's [1968, 1971, 1975] theory of bureaucratic preferences or a modified extension of the Capture Theory to allow for competition in the political process along the lines of Becker [1983].

Once the pure form of CT regulation is relaxed it is likely that the clear dichotomy between the optima of the CT and NPIT regulators would become blurred. This was evident in the brief consideration of the regulation of a monopoly where the concept of strategic regulation was discussed. There the regulatory process was a form of game in which a CT regulator would tone down its activities in order to disguise its motives and thus preclude more hostile NPIT regulation. A similar process would occur in a Becker-type model. Although results would inevitably be less clear-cut, in principle the ability to infer a regulator's objectives from observable actions remains albeit in terms of a probability-weighted determination from a continuum of possible regulatory behaviour.

In the models developed here, the choice variables of the regulator were the level of enforcement and, in part, the regulatory instrument. It was evident from the analysis, however, that optimal regulatory policy was much affected by the form of the expected penalty function. Accordingly, another obvious extension would be to allow the regulator the choice of the structure and amount of the expected penalty. This would likely provide further observable characteristics by which the objectives of the regulator could be inferred.

The general equilibrium model developed in the thesis was relatively simple in structure. A more general treatment of an economy containing many industries and producing many commodities would also be a useful exercise. It is doubtful, however, whether many of the results would survive such a generalization. The robustness and variety of the initial results presented here were due to the strength and specificity of the assumptions used. Even with the relatively slight degree of generalization involved in the development of the simple general equilibrium model it was evident that the core of unqualified results was somewhat reduced.

Finally, some degree of empirical investigation is required to examine the applicability and validity of the theory in practice. This might best be done in the first instance with reference to well-documented cases of past regulation for which data is perhaps more readily available and the predictive performance of the model can be tested against known outcomes. Regrettably, the scope of such a project was beyond the time and resources available to the present study. Once any refinements that proved necessary in response to this exercise were incorporated into the structure of the model and its validity determined, a possible use to the policy maker would be as a form of "acid test" to identify regulators which appeared to be operating in accordance with objectives other than those of society as a whole, and upon whom more detailed scrutiny would be warranted.

# ACKOWLEDGEMENTS

There are a number of people I wish to thank for their help and support in the writing and production of this thesis.

First and foremost is my supervisor Richard Manning, now of State University of New York at Buffalo,who encouraged me to embark upon this project,inspired me in its execution and provided tireless expert assistance and guidance throughout its duration. Ken Henry,my associate supervisor,and Alan (Bruce) Woodfield,quasi-associate supervisor,also proffered much valuable advice during my supervisor's sabbatical. They and the other "Bruces" of the Canterbury Economics Department, most notably Frank Tay,Ewan M$^{c}$Cann,Brent Layton and Tony Raynor,will always remain fondly in my thoughts as the teachers,colleagues and friends who introduced me to the fascinating realm of economics.

My thoughts and best wishes also remain with my fellow doctoral students Paul Dalziel,Terri Green,Seamus Hogan and Ivette Jans who were a source of great encouragement and much camaraderie during our time together at Canterbury.

I am most grateful too to the Reserve Bank of New Zealand and the University Grants Committee for their generous financial support and to Karilyn Smith for her skilful and patient typing of what must have seemed an eternity of tedious mathematical expressions.

Finally I would like to thank my parents for all the opportunities they so generously provided me with and especially my wife Fredericka,to whom I dedicate this thesis,who through her patience and love,and that of our family as well, lent me the emotional support necessary to persevere in my moments of despondency.

REFERENCES

ANDERSON, L.G. (1976) The relationship between firm and industry in common property fisheries. Land Economics, 52(2), May; 179-191.

ANDERSON, L.G. (1977) The economics of fisheries management. Baltimore, John Hopkins University Press.

AVERCH, H. and JOHNSON, L.L. (1962) Behaviour of the firm under regulatory constraint. American Economic Review, 52(5), December; 1052-1069.

BARON, D.P. and MYERSON, R.B. (1982) Regulating a monopolist with unknown costs. Econometrica, 50(4), July; 911-930.

BAUMOL, W.J. and OATES, W.E. (1981) Pollution control: direct control versus the pricing system. In Wagner, L. (ed) Readings in applied microeconomics. 2nd ed. Oxford, Oxford University Press. p.353-368.

BECKER, G.S. (1968) Crime and punishment: an economic approach. Journal of Political Economy, 76 (March/April); 169-217.

BECKER, G.S. (1976) Comment on "Peltzman: Toward a more general theory of regulation". Journal of Law and Economics, 19(2), August; 245-248.

BECKER, G.S. (1983) A theory of competition among pressure groups for political influence. Quarterly Journal of Economics, 98(3), August; 371-400.

BECKER, G.S. and STIGLER, G.J. (1975) Law enforcement, malfeasance and compensation of enforcers. In Selden, R.T. (ed). Capitalism and freedom: problems and prospects. Charlottesville, University Press of Virginia. p.230-261.

BERNSTEIN, M.H. (1977) Regulating business by independent commission. Connecticut, Greenwood Press. (First published in 1955; Princeton, N.J.; Princeton University Press.)

BHAGWATI, J.N. (1982) Directly unproductive profit seeking (DUP) activities. Journal of Political Economy, 90, October; 988-1002.

BLOCK, M.K. and LIND, R.C. (1975) Crime and punishment reconsidered. Journal of Legal Studies, 4(1), January; 241-247.

BROOKS, M.A. and HEIJDRA, B.J. (1986) In search of rent seeking. University of Tasmania (Economics Discussion Papers. DP 86-01).

BUCHANAN, J.M. and TULLOCK, G. (1975) Polluters' profits and political response: direct controls versus taxes. American Economic Review, 65(1), March; 139-147.

CARR-HILL, R.A. and STERN, N.H. (1975) Crime, the police and criminal statistics. London; Academic Press.

CHEUNG, S. (1970) The structure of a contract and the theory of a non-exclusive resource. Journal of Law and Economics, 13(1), April; 49-70.

CHRISTY, F.T. (Jnr) and SCOTT, A. (1965) The common wealth in ocean fisheries: some problems of growth and economic allocation. Baltimore, Maryland; John Hopkins Press.

CLARK, C.W. (1982) Models of fishery regulation. In Mirman, L.J. and Spulber, D.E. (eds) Essays in the economics of renewable resources. Amsterdam; North Holland. p.273-287.

CLARK, C.W. and MUNRO, G.R. (1982) The economics of fishing and modern capital theory. In Mirman, L.J. and Spulber, D.W. (eds) Essays in the economics of renewable resources. Amsterdam; North Holland. p.31-54.

CLONINGER, D.O. (1975) The deterrence effect of law enforcement: an evaluation of recent findings and some new evidence. American Journal of Economics and Sociology, 34(3), July; 323-335.

COASE, R.H. (1960) The problem of social cost. Journal of Law and Economics, 3, October; 1-44.

COPES, P. (1970) The backward-bending supply curve of the fishing industry. Scottish Journal of Political Economy, 17(1), February; 19-77.

COPES, P. (1972) Factor rents, sole ownership and the optimum level of fisheries exploitation. Manchester School of Economics and Social Studies, 40(2), June; 145-164.

DAHLMAN, C. (1980) The open field system and beyond: a property rights analysis of an economic institution. Cambridge; Cambridge University Press.

DIVER, C.S. (1980) A theory of regulatory enforcement. Public Policy, 28(3), Summer; 257-299.

DNES , A.W. (1985) Rent seeking and open-access fishing. Scottish Journal of Political Economy, 32(2), June; 159-170.

ECKERT, R.D. (1973) On the incentives of regulators: the case of taxicabs. Public Choice, 14, Spring; 83-100.

ERHLICH, I. (1972) The deterrent effect of law enforcement. Journal of Legal Studies, 1(2), June; 259-276.

EHRLICH, I. (1973) Participation in illegal activities: a theoretical and empirical investigation. Journal of Political Economy, 81(3); May/June; 521-565.

EHRLICH, I. (1982) The optimum enforcement of laws and the concept of justice: a positive analysis. International Review of Law and Economics, 2; 3-27.

EPPLE, D. and VISSCHER, M. (1984) Environmental pollution: modelling occurrence, detection and deterrence. Journal of Law and Economics, 27(1), April; 29-60.

FRIEDMAN, M. (1962) Capitalism and freedom. Chicago; University of Chicago Press.

FRIEDMAN, M. and SCHWARTZ, A.J. (1963) A monetary history of the United States, 1867-1960. Princeton, N.J.; Princeton University Press.

FULLERTON, D. (1982) On the possibility of an inverse relationship between tax rates and government revenues. Journal of Public Economics, 19(1), October; 3-22.

GOLDBERG, V.P. (1976) Regulation and administered contracts. Bell Journal of Economics, 7, Autumn; 426-448.

GORDON, H.S. (1954) The economic theory of a common property resource: the fishery. Journal of Political Economy, 62; 124-142.

HARDIN, G. (1968) The tragedy of the commons. Science, 162; 1243-1248.

HARFORD, J.D. (1978) Firm behaviour under imperfectly enforceable pollution standards and taxes. Journal of Environmental Economics and Management, 5(1), March; 26-43.

HAUSMAN, J.A. (1981) Exact consumer surplus and dead-weight loss. American Economic Review, 71(4), September; 662-676.

HAWKINS, K. and THOMAS, J.M. (1984) The enforcement process in regulatory bureaucracies. In Hawkins, K. and Thomas, J.M. (eds) Enforcing regulation. Boston; Kluwer-Nijhoff Publishing. p.3-22.

HIRSCHLIEFER, J. (1976) Comment on "Peltzman: Toward a more general theory of regulation". Journal of Law and Economics, 19(2), August; 241-244.

KAGAN, R.A. and SCHOLZ, J.T. (1984) The "criminology of the corporation" and regulatory enforcement strategies. In Hawkins, K. and Thomas, J.M. (eds) Enforcing regulation. Boston; Kluwer-Nijhoff Publishing. p.67-95.

KEMP, M. and NG, W. (1979) The importance of being honest. Economic Record, 55, March; 41-46.

KOLKO, G. (1963) The triumph of conservatism: a reinterpretation of American history, 1900-1916. London; The Free Press of Glencoe Collier-Macmillan Ltd.

KOSTERS, M.H. (1979) Counting the costs: the Business Roundtable study. Regulation, 3 , July/August ; 17-25.

LEE, D.R. (1984) The economics of enforcing pollution taxation. Journal of Environmental Economics and Management, 11; 147-160.

LINDER, S.H. and McBRIDE, M.E. (1984) Enforcement costs and regulatory reform: the agency and firm response. Journal of Environmental Economics and Management, 11; 327-346.

LOEB, M. and MAGAT, W.A. (1979) A decentralized method for utility regulation. Journal of Law and Economics, 22(2), October; 399-404.

MALONEY, M.T. and McCORMICK, R.E. (1982) A positive theory of environmental quality regulation. Journal of Law and Economics, 25(1), April; 99-123.

MANNING, R. (1986a) Consumer choice; chapter three of lecture notes for ECON 301/351, Allocation theory. Department of Economics and Operations Research. University of Canterbury (unpublished).

MANNING, R. (1986b) Ownership concentration and the efficiency of monopoly. University of Canterbury, Department of Economics and Operations Research (Discussion Paper 8610).

MATHUR, V.K. (1978) Economics of crime: an investigation of the deterrent hypothesis for urban areas. Review of Economics and Statistics, 60(3), August; 459-466.

MIGUÉ, J.L. (1977) Controls versus subsidies in the economic theory of regulation. Journal of Law and Economics, 20(1), April; 213-221.

MIRMAN, L.J. and SPULBER, D.F. (1985) Fishery regulation with harvest uncertainty. International Economic Review, 26(3), October; 731-746.

MONTGOMERY, W.D. (1972) Markets in licenses and efficient pollution control programs. Journal of Economic Theory, 5(3), December; 395-418.

McKEAN, R.N. (1980) Enforcement costs in environmental and safety regulation. Policy Analysis, 6(3); 269-289.

NISKANEN, W.A. (1968) Nonmarket decision making and the peculiar economics of bureaucracy. American Economic Review, 58, May; 293-305.

NISKANEN, W.A. (1971) Bureaucracy and representative government. Chicago, Aldine-Atherton.

NISKANEN, W.A. (1975) Bureaucrats and politicians. Journal of Law and Economics, 18(3), December; 617-643.

OWEN, B.M. and BRAEUTIGAM, R. (1978) The regulation game: strategic use of the administrative process. Cambridge, Mass., Ballinger Publishing Company.

PAPPS, I. (1985) Controlling and monitoring pollution. A comment on "The criminal waste discharger". Scottish Journal of Political Economy, 32(2), June; 171-180.

PELTZMAN, S. (1976) Toward a more general theory of regulation. Journal of Law and Economics, 19(2), August; 211-240.

PELTZMAN, S. (1977) The effect of regulation on new drugs. In Siegan, B.H. (ed) The interaction of economics and the law. Lexington, Lexington Books, D.C. Heath and Company. p.101-111.

POLINSKY, A.M. (1979) Controlling externalities and protecting entitlements: property right, liability rule, and tax-subsidy approaches. Journal of Legal Studies, 8(1), January; 1-48.

POLINSKY, A.M. and SHAVELL, S. (1979) The optimal tradeoff between the probability and magnitude of fines. American Economic Review, 69(5), December; 880-891.

POSNER, R.A. (1969) Monopoly and regulation. Stanford Law Review, 21, February; 548-643.

POSNER, R.A. (1974) Theories of economic regulation. Bell Journal of Economics and Management Science, 5(2), Autumn; 335-358.

POSNER, R.A. (1975) The social costs of monopoly and regulation. Journal of Political Economy, 83(4), August; 807-827.

PUTSAY, M.W. (1978) The social costs of monopoly and regulation: an empirical investigation. Southern Economic Journal, 45(2), October; 583-591.

PYLE, D.J. (1983) The economics of crime and law enforcement. London, Macmillan Press.

QUIRK, J.P. and SMITH, V.L. (1970) Dynamic economic models of fishing. In Scott, A.D. (ed) Economics of fisheries management: a symposium. Institute of Animal Resource Ecology. University of British Columbia.

ROSE, N.L. (1985) The incidence of regulatory rents in the motor carrier industry. The Rand Journal of Economics, 16(3), Autumn; 299-318.

ROSS, T.W. (1984) Uncovering regulators' social welfare weights. The Rand Journal of Economics, 15(1), Spring; 152-155.

ROSS, T.W. (1985) Extracting regulators' implied welfare weights: some further developments and applications. Quarterly Review of Economics and Business, 25(3), Autumn; 72-84.

RUNGE, C.F. (1981) Common property externalities: isolation, assurance and resource depletion in a traditional grazing context. American Journal of Agricultural Economics, 63(4), November; 595-606.

SALOP, S.C. and SCHEFMAN, D.T. (1983) Raising rivals' costs. American Economic Review: Papers and Proceedings, 73(2), May; 267-271.

SCOTT, A. (1955) The fishery: the objectives of sole ownership. Journal of Political Economy, 63(2), April; 116-124.

SCOTT, A. and SOUTHEY, C. (1970). The problem of achieving efficient regulation of a fishery. In Scott, A.D. (ed) Economics of fisheries management: a symposium. Institute of Animal Resource Ecology. University of British Columbia.

SIMON, M.E. (1981) Government regulation: adding up the cost. Journal of Contemporary Business, 9(2); 5-16.

SMITH, R.T. (1976) The legal and illegal markets for taxed goods: pure theory and an application to the state government taxation of distilled spirits. Journal of Law and Economics, 19(2), August; 393-429.

SMITH, V.L. (1966) Review of Christy and Scott, "The common wealth in ocean fisheries". American Economic Review, 56(5), December; 1341-1343.

SMITH, V.L. (1968) Economics of production from natural resources. American Economic Review, 58(3), June; 409-431.

SMITH, V.L. (1969) On models of commercial fishing. Journal of Political Economy, 77(6), January/February; 181-198.

STIGLER, G.J. (1970) The optimum enforcement of laws. Journal of Political Economy, 78(3), May/June; 526-536.

STIGLER, G.J. (1971) The theory of economic regulation. Bell Journal of Economics and Management Science, 2(1); 3-21.

STIGLER, G.J. (1975) The citizen and the state: essays on regulation. Chicago, University of Chicago Press.

STIGLER, G.J. (1981) Comment on Joskow, P.L. and Noll, R.C., "Regulation in theory and practice: an overview." In Fromm, G. (ed) Studies in public regulation. Cambridge, Mass., MIT Press. p.73-78.

STIGLER, G.J. and FRIEDLAND, C. (1962) What can regulators regulate? the case of electricity. Journal of Law and Economics, 5(3); 1-16.

STOREY, D.J. and McCABE, P.J. (1980) The criminal waste discharger. Scottish Journal of Political Economy, 27(1); 30-40.

TOLLISON, R.D. (1982) Rent seeking: a survey. Kyklos, 35(4); 575-602.

TULLOCK, G. (1967) The welfare costs of tariffs, monopolies, and theft. Western Economic Journal, 5, June; 224-232.

TURVEY, R. (1964) Optimization and suboptimization in fishery regulation. American Economic Review, 54(1), March; 64-76.

VELJANOWSKI, C.G. (1983) The market for regulatory enforcement. In Supplement to the Economic Journal. Conference papers from the annual conference of the Royal Economic Society. Cambridge, Cambridge University Press. p.123-129.

VELJANOWSKI, C.G. (1984) The economics of regulatory enforcement. In Hawkins, K. and Thomas, J.M. (eds) Enforcing regulation. Boston, Kluwer-Nijhoff Publishing. p.171-188.

VISCUSI, W.K. and ZECKHAUSER, R.J. (1979) Optimal standards with incomplete enforcement. Public Policy, 27(4), Fall; 437-456.

WAUD, R.N. (1986) Tax aversion and the Laffer curve. Scottish Journal of Political Economy, 33(3), August; 213-227.

WEIDENBAUM, M.L. (1980) Costs of regulation and benefits of reform. Centre for the Study of American Business (Publication 35).

WEITZMAN, M.L. (1974) Prices vs quantities. Review of Economic Studies, 41(4), October; 477-492.

WEITZMAN, M.L. (1978) Optimal rewards for economic regulation. American Economic Review, 68(4), September, 683-691.

WILLIG, R. (1976) Consumer surplus without apology. American Economic Review, 66, September; 589-597.

# AN ECONOMIC ANALYSIS OF
# REGULATORY OBJECTIVES AND ENFORCEMENT
## ADDENDUM[1]

This addendum provides further interpretation and clarification of several issues considered in Chapters Two and Three of the accompanying thesis.[2]

(1) <u>Formulation of the model</u>

Expression (2-4-8) is not meant as a prior constraint on the model. Rather it is a conclusion from the model which follows the assumptions on the expected penalty function concerning full reporting and over-reporting.

Expression (2-4-7) includes a component $G(k,k) = 0$; that is, no penalty is incurred for truthful declaration. This coupled with the assumption that no penalty is incurred for over-reporting implies that

(A-1) $$G(k,n) = 0 \quad \forall\ n > k$$

and therefore

(A-2) $$G_x = 0 \quad \forall\ x > q \text{ and } \quad G_q = 0 \quad \forall\ q < x$$

Using (A-1) and (A-2) it then immediately follows from (2-4-14) that $x^* \leq q^*$ because for any $x > q$ $G_x = 0$ and $\partial\pi/\partial x < 0$. An additional unit of declared output that is not actually produced merely results in additional tax liability which cannot be optimal.

The maximization of $\pi(q,x,\alpha)$ shown in (2-4-11) is therefore unconstrained and the "greater than" inequality shown in (2-4-14) should be deleted as it should also in (3-4-18). It is possible, however, to conceive of an enforcement structure which would ensure truthful behaviour by the firm. Certainty of detection coupled with a unit rate penalty on undeclared output which exceeds the tax rate is one such structure.

In this case

(A-3) $\pi_x > 0 \quad \forall\ x < q$ and $\pi_x < 0 \quad \forall\ x > q$

For such a penalty structure in the limit as x approaches q from below profit can be increased by more truthfully declaring actual output resulting in an optimum when output is fully declared.

(2) Form of the expected penalty function

One intuitively appealing form of the expected penalty function is that the expected penalty is related to the amount of undeclared output. That is

(A-4) $G = G(q-x)$

This form is used extensively in the thesis. The word "extent" as used in Sections 2-4 through 2-6 and subsequent analysis based on these sections should be interpreted as "amount".

Other interesting forms of the expected penalty function might also have been considered. It may, for instance, be thought reasonable for the expected penalty to be related to the proportion of undeclared output. Alternatively, illegal behaviour by large firms might be thought worthy of more stringent punishment than if the firms were small. An argument could also be based on the significance or relative size of an industry in the economy as a whole. Under these different definitions of the expected penalty the validity of the propositions of this thesis would have to be considered on a case by case basis.

Expression (2-4-26) when correctly evaluated at q shows the penalty-inclusive marginal cost of producing an additional unit of undeclared output. The form of (2-4-27) and the subsequent analysis based on it contained in the proofs of Propositions 2-4-2, 2-4-5 and 2-4-8 are correct under the condition that

(A-5) $\quad G_q = - G_x$

One such case occurs with an expected penalty function of the form of (A-4). The condition shown in (A-5) need not hold in general, however. The general formulation of (2-4-27) is

(A-6) $\quad MC(q) = c'(q) + t^0 + G_q + G_x$

Under the general formulation of the marginal cost of producing a unit of declared output shown in (A-6) the alternative proofs of Propositions 2-4-5 and 2-4-8 given in the text of this thesis are justified. A similar proof drawing on (2-4-14) and the assumption of convexity of the expected penalty function can be used to demonstrate the validity of Proposition 2-4-2 as follows:

> Comparing (A-6) and (2-4-26) any output produced is undeclared for all q such that $G_x > - t^0$, a condition which follows immediately from (2-4-14). Following (2-4-25) this condition holds for at least the initial unit of output. From Proposition 2-4-1 $G_{xq} < 0$ and an increase in output increases the absolute value of the marginal expected penalty with respect to declared output. This output-induced increase in $G_x$ may eventually be sufficient to equate its absolute value to the tax rate. Given that supply rises with price, at all prices below that necessary to call forth the output level at which the absolute value of $G_x$ is equated with the tax rate any output produced will be undeclared.

The proof of Proposition 2-5-1 together with the form of Figure 2-5-1 and the equilibrium analysis which follows from it is also based on an expected penalty of form such that (A-5) holds. Relaxation of this condition in other cases admits the possibility of several additional types of equilibria beyond those considered in Section 2-5.

One such possibility, using the terminology of Figure 2-5-1, is an equilibrium $Q'$ where

(A-7) $\quad Q_t^* < Q' < Q^0$

such that

(A-8) $\quad |G_{X'}| = t > G_{Q'} \;;\; X' < Q'$

(A-7) and (A-8) together denote an equilibrium $(Q',X')$ with some declared output at an output level intermediate between that associated with the unregulated competitive equilibrium and that associated with full declaration.

Another possibility is an equilibrium $Q''$ where

(A-9) $\quad Q'' < Q_t^*$

such that either

(A-10) $\quad |G_{X''}| < t < G_{Q''}$

in which case $X'' = 0$, or

(A-11) $\quad |G_{X''}| = t < G_{Q''} \;;\; X'' < Q''$

(A-9) and (A-10) together denote an equilibrium $(Q'',0)$ with no declared output at an output level below that associated with full declaration while (A-9) and (A-11) denote an equilibrium $(Q'',X'')$ with partial declaration of an output level below that associated with full declaration.

It is possible, therefore, to have an expected penalty function such that the marginal expected penalty with respect to output given that no output is declared does not exceed the tax rate yet the equilibrium involves partially declared output and it is also possible to have an expected penalty function such that the marginal expected penalty with respect to output at the equilibrium exceeds the tax rate yet the equilibrium involves either zero or partial declaration. In the first case the regulator is more concerned with encouraging truthful behaviour than restricting output while in the latter case the reverse is true. For the purposes of this addendum no further analysis of the properties of these possible equilibria is undertaken.

(3) Redistribution of fine revenue

The analysis of CT regulation in Sections 3-4 and 3-5 contains a term 'a' which represents the proportion of fine revenue redistributed to members of the regulated industry. The intention of the model is that any redistribution of such proceeds occurs in some lump-sum manner and hence marginal decisions of firms are unaffected by it. It may be understood therefore that (3-4-2) and (3-4-4) represent the net expected penalty received by the regulator but not that upon which individual firms base their output decisions. It may also be considered that in most situations 'a' would be zero.

If, however, 'a' was known in advance to individual firms and proceeds were redistributed in other than a lump-sum manner the marginal behaviour of firms and hence the regulated equilibrium would be affected. In such a case the form of terms J in (3-5-1) and U in (3-5-2) would be similar to that of terms E and S in (3-5-1) and (3-5-2) repsectively and the form and proof of Proposition 3-5-3 would be similar to those of Proposition 3-5-4.

## NOTES

1 Between the initial completion of this thesis and the writing of this addendum Richard Manning tragically and prematurely passed away. This thesis is but one humble testimony to his enduring contribution to economic thought in New Zealand and beyond.

2. The author is indebted to Professor Lewis Evans of Victoria University of Wellington and Alan Woodfield of the University of Canterbury for helpful discussions on the matters considered in this addendum.